DEPARTMENT OF HEALTH AND HUMAN SERVICES
CENTERS FOR DISEASE CONTROL AND PREVENTION

NATIONAL PROGRAM OF CANCER REGISTRIES

TECHNICAL ASSISTANCE AND AUDIT

ARIZONA CANCER REGISTRY

Case Completeness and Data Quality Audit

September 2003

Prepared by

ORC MACRO℠

This program is funded through the Centers for Disease Control and Prevention
Contract Number: 200-2000-10059

ACKNOWLEDGMENTS

The staff of the Technical Assistance and Audit Program would like to thank the director of the Arizona Cancer Registry, Georgia Armenta Yee; and her staff; and the staff of the participating hospitals for their patience and their indefatigable assistance with this audit. Its success was made possible by their generous and diligent efforts. We gratefully acknowledge their invaluable contribution to achieving standards of data quality and completeness in the collection of data on cancer.

We also thank the Centers for Disease Control and Prevention (CDC) project officers—Faye Floyd, Leah Simpson, and Kimberly Cantrell—for their participation in and guidance on the technical and administrative aspects of the audit.

Last, but not least, we thank the ORC Macro Publications Services Department for its contributions in preparing this report.

Page

APPENDICES

Tables and Charts

I. INTRODUCTION

Since the beginning of the National Program of Cancer Registries (NPCR), the Division of Cancer Prevention and Control within the Centers for Disease Control and Prevention (CDC) has provided States funded under NPCR with assistance in developing and enhancing State cancer registries, as well as with effective registry operations and the monitoring of completeness, timeliness, and quality of data under the auspices of Public Law 102-515 (the National Cancer Registries Amendment Act). States are responsible for ensuring compliance with NPCR program standards for completeness, timeliness, and the quality of data reported to the central cancer registry.

From 1995 through 2000, with funding from CDC, the Public Health Institute (PHI) conducted audits of data completeness and quality through the Cancer Surveillance and Control Program (CSCP). The North American Association of Central Cancer Registries (NAACCR) acted as an advisor to PHI in the design, maintenance, and assessment of CSCP. A total of 38 States was audited at least once during that period.

As of October 1, 2000, Macro International Inc. (ORC Macro) was awarded a 5-year contract to assess the completeness and accuracy of data from the central cancer registry in the States and to provide technical assistance in their operations under the auspices of NPCR. The Technical Assistance and Audit (TAA) Program follows the guidelines set by CDC and NAACCR in assessing the completeness and quality of data collected by State cancer registries, providing a comparison of the State's performance with NPCR States' average and recommending approaches that could result in improvement of the State cancer registry's data completeness and accuracy. TAA completed 17 State audits during the first 2 years of the contract.

II. PURPOSE

The primary purpose of the NPCR Case Completeness and Data Quality Audit is to assess the case completeness and the level of quality of data collected by NPCR-funded, statewide, population-based cancer registries. These data are a crucial part of cancer surveillance systems because they are used for planning, operating, funding, and evaluating cancer control programs. Complete and accurate data are essential for estimating variations in population subgroups and changes among population subgroups over time. The audit assessment is based on the existence of the following:

1) Appropriate policies and procedures for data collection

2) Appropriate policies and procedures for assessment of data quality

3) Data related to female breast, colon and rectum, lung and bronchus, and prostate cancers.

III. CONFIDENTIALITY AND SECURITY

All audit functions are performed under the pertinent confidentiality statutes. TAA staff signed the necessary confidentiality agreements before they were given access to confidential material. Confidential data accessed by TAA auditors during the audit were used only for the purpose of conducting the audit. Upon completion of the audit's data collection, analysis, and reporting activities, confidential data used during the audit process were either returned to the State or destroyed as required by the statement of disposition in the confidentiality agreement.

IV. MATERIALS AND METHODS

ELIGIBILITY FOR NPCR AUDIT

All States receiving funding from CDC for the operation of a central cancer registry are eligible for an NPCR audit. States that have not been audited or that were audited more than 3 years ago may be given priority in the selection process.

Military and Veterans Administration hospitals are not included in this audit because they are not subject to State laws. Children's hospitals are also not included in this audit. All other hospitals that report more than 24 new cancer cases and that are required by Arizona State law to report cancer cases to the Arizona Cancer Registry (ACR) are eligible for participation in this audit.

DATA SOURCES

ACR prepared an extract file, a "master abstract file," of all in situ and invasive cancer cases of female breast, colon and rectum, lung and bronchus, and prostate that were diagnosed in 2000. That file was an unduplicated file—that is, it did not contain multiple facility reports for the same reportable malignancy. ACR also provided a master hospital list with the number of analytic cases (class of case 0, 1, or 2) of the four cancer sites for each hospital. The list ensured that the hospital was placed into the proper caseload category.

TAA staff used PC-SAS software for the sampling. The PC-based NPCR-TAA utility was used for record matching, statistical analysis, and report production. That utility, on laptop computers, was also used by TAA auditors for record reabstraction and casefinding. Data completeness and accuracy rates were computed as mentioned in the sections on Casefinding and Data Quality.

By means of statistical evaluation, each facility was placed into a caseload category. The three categories were based on the total analytic cases of female breast, colon and rectum, lung and bronchus, and prostate cancer the hospitals reported in 2000. The hospital(s) with the highest number of cases reported and the hospital(s) with the lowest number of cases reported provided the frame. The remaining hospitals were then divided equally between caseload categories on the basis of their caseload and therefore fell into one of the following three categories:

High caseload:	409 to 613 cases
Medium caseload:	204 to 408 cases
Low caseload:	25 to 203 cases

For casefinding, 6 months of records were examined in the low-caseload facilities, 4 months of records were examined in the medium-caseload facilities, and 3 months of records were examined in the high-caseload facilities.

SELECTING HOSPITAL SAMPLES

Sample hospitals are selected using the probabilities proportional to size (PPS) method. The basic concept of PPS sampling is that the probability of selecting a hospital is proportional to its size. Therefore, when the PPS methodology is used to determine a sample, a hospital with 300 cases would have twice the probability of being selected that a hospital with 150 cases does.

If 10 hospitals are to be selected in a State with 2,000 cancer cases, the probability of selection for a hospital with 200 cases would be 1.0 (i.e., 200*10/2000 = 1) (see appendix—Calculating Variances and Standard Errors).

The chance of a hospital's being selected increases as the caseload of the hospital increases. For example, a hospital with 400 cases has a chance of being selected twice (400*10/2000 = 2).

If a facility is selected more than once by the PPS model, samples for data quality are also increased accordingly. For example, in the hospital with 400 cases, record samples are doubled.

SAMPLING FOR MEASURING DATA QUALITY

After a hospital was selected, a fixed number of cases (33 records) was selected by means of a simple-random-sampling model. The selection of a fixed number of cases with equal probabilities, coupled with PPS selection of hospitals, yields an overall sample with equal probabilities—that is, an approximately self-weighting sample. Self-weighting samples facilitate the analysis and enhance the statistical efficiency of sample estimates.

A total of 297 cases was reviewed to assess the data quality in sample facilities. This sample size was determined to achieve sufficient precision for an expected case completion rate of 95 percent.

CASEFINDING PROCESS

The level of case completeness during a selected period is assessed by independently casefinding cancer cases in sample facilities. TAA auditors use an electronic audit program customized for this audit. All sources in each of the hospitals were audited for case completeness, including the following:

- Medical Record Disease Indices (MRDI)
- Pathology reports (including bone-marrow, autopsy, and other specialized pathology reports)
- Cytology reports
- Surgical logbooks and same-day-surgery logbooks
- Outpatient clinic records
- Radiation therapy (RT) clinic logs

- Nuclear-medicine logs
- Any other source in the hospital where patients with female breast, colon and rectum, lung and bronchus, and prostate cancer were diagnosed and/or treated.

Reportable cancers that are included in the casefinding activities were determined after reviewing the reporting practices—including reporting requirements, procedure manuals, and coding practices of the State.

Any new incident cancer case for the audit year that was not in the master extract file was considered a missed case. Completeness rates were computed for each caseload category. These rates were applied to the proportion of incident cancer cases in the caseload category for the State.

Case Completeness Rate (%) =100 – (Number of Missed Cases x 100/Total Number of Cases Identified)

QUALITY ASSURANCE PROCESS

Reabstracting audits are done to assess the accuracy (agreement with source medical records) and reproducibility (agreement among data collectors) of registry data. The purposes of reabstracting and recoding studies are as follows:

- To standardize interpretation and abstracting of the medical record
- To estimate rates of agreement
- To identify problems in data collection and interpretation.

TAA auditors reabstracted and recoded data from the source records (in most cases, the hospital medical record) and compared the codes with the data already in the registry to determine whether the codes matched exactly. Reabstracted cancer cases became the standard against which the previously abstracted cases already in the cancer registry were compared. Because the auditors were reviewing a medical record against a "case"—which might contain consolidated information from several other sources—precautions were taken so as not to assess the accuracy of the underlying medical record when attempting to measure the reproducibility of data collection and coding.

Cases with unmatched codes were returned to the registry for reconciliation with information from additional sources. A code discrepancy was determined if the consensus code did not exactly match the registry's original code.

Data accuracy rate estimates were computed for each caseload category. These rates were applied to the proportion of incident cancer cases in the caseload category for the State.

*Error Rate (%) = Number of Discrepancies x 100/Total Data Elements**

** Total Data Elements = Number of Records Reabstracted x Number of Data Elements Reviewed*

V. AUDIT WORK PLAN

MASTER EXTRACT FILE OF ALL REPORTABLE CASES OF FEMALE BREAST, COLON AND RECTUM, LUNG AND BRONCHUS, AND PROSTATE CANCER

ACR prepared a master extract file of cancer cases diagnosed in 2000 among Arizona residents. This master extract file contained consolidated records of multiple abstracts from different facilities for the same reportable cancer for the above-mentioned sites. The master extract file was submitted in NAACCR 2000 record layout, version 9.

ACR provided a master hospital list of all hospitals in the State, as well as Veterans Administration hospitals and children's hospitals. The number of analytic cases reported by each of the facilities was included.

A total of 10,377 eligible cases of female breast, colon and rectum, lung and bronchus, and prostate cancer was included in the master extract file submitted by ACR for the diagnosis year 2000. Of those, 3,413 were female breast, 2,370 were colon and rectum, 2,634 were lung and bronchus, and 1,960 were prostate cancer cases. Those cases were reported from 54 hospitals in Arizona. 13 hospitals with fewer than 25 reported cases, Federal hospitals, and children's hospitals were omitted from the sampling frame.

HOSPITAL AND CASE SAMPLES

A total of 9 hospitals was selected by means of PPS modeling. Hospitals selected by caseload and the number of months audited were as follows:

Number of Hospitals Selected	Caseload	Number of Months Audited
4	High	3
3	Medium	4
2	Low	6

REABSTRACTING ACTIVITIES

Two TAA auditors visited nine hospitals chosen in this random sample. The sample provided an unbiased, independent assessment of the quality of the data because the TAA auditors had not previously been involved in the reporting or abstracting of any of the cases in the audit sample.

One TAA auditor was responsible for the reabstraction portion of the audit. The data were reabstracted into the NPCR-TAA audit utility that had been customized for this audit. The master file was uploaded into the NPCR-TAA utility on laptop computers used by the TAA auditors.

To assess the quality of data, the auditors reabstracted the following data elements:

1) Demographic information, including

    a. *State of Residence at Diagnosis*

    b. *Race*

    c. *Sex*

    d. *Date of Birth (mm/dd/yyyy).*

2) Pathology data characterizing the cancer, including

    a. *Date of Diagnosis (mm/dd/yyyy)*

    b. *Primary Site (first three digits of the International Classification of Diseases for Oncology [ICD-O] topography code)*

    c. *Subsite (fourth digit of the ICD-O topography code)*

    d. *Morphology (first four digits of the ICD-O morphology code)*

    e. *Behavior (fifth digit of the ICD-O morphology code)*

    f. *Grade (sixth digit of the ICD-O morphology code)*

    g. *Sequence Number*

    h. *Laterality*

    i. *Stage at Diagnosis (SEER Summary Stage).*

Each day when the reabstraction audit was completed for a facility, the database was saved in a password-protected Zip file and downloaded on a computer diskette to ensure security and confidentiality and to serve as a backup. Likewise, each day the TAA principal investigator, program manager, and statistician monitored the audit team's progress to identify potential problems in the audit process and resolve any issues with the State program personnel.

CASEFINDING ACTIVITIES

One TAA auditor performed the casefinding audit. The following sources in each hospital were reviewed:

1) Medical Record Disease Indices (MRDI)

2) Pathology reports (including autopsies, bone-marrow, and other specialized pathology reports)

3) Cytology reports (if separate from pathology reports)

4) Radiation therapy clinic logs

5) Surgical logbooks and same-day-surgery logbooks

6) Outpatient-clinic records

7) Nuclear-medicine logs

8) Any other source in the hospital where patients with female breast, colon and rectum, lung and bronchus, or prostate cancers were diagnosed and/or treated.

When a reportable cancer for any of the audited sites (female breast, colon and rectum, lung and bronchus, or prostate) within the specified diagnosis year was found in any of the above-mentioned casefinding sources that did not match the master extract file of reported cases, a new case accession was created and added to the casefinding audit database. The casefinding audit database contained all the cases that the auditors considered "missed."

At the end of each workday, the casefinding audit database was saved in a password-protected ZIP file on a computer diskette to ensure the security and confidentiality of the data and to serve as a backup.

Upon completion of the onsite audit, the director and staff of the central cancer registry were briefed and given an overview of the audit, with preliminary observations.

VI. RECONCILIATION

TAA staff analyzed the missed records and data discrepancies. Both files were matched against the master extract file submitted by ACR. Lists and printed abstracts of unmatched cases were prepared and returned to ACR for reconciliation within a week of the onsite audit. ACR staff collaborated with hospital registry staff to reconcile records that did not match the master extract file—that is, "possible missed cases" or "possible errors." Possible missed cases may be the result of information on non-Arizona residents, nonreportable cancers, or diagnosis years before 2000. Possible errors in data quality may be the result of either more appropriate information found from a different facility or problems with record consolidation. The lists were returned to the TAA team on completion of the reconciliation.

Possible missed cases and data elements with possible errors that could not be resolved after the reconciliation process were considered "missed cases" and "errors."

VII. RESULTS AND DISCUSSION

CASEFINDING

The number of cases identified as "missed" and the sources from which those cases were identified are shown in table 1A. There was a total of 21 missed cases. Of the missed cases, 20 (95.2 percent) were identified in one casefinding source. One case (4.8 percent) was identified in two casefinding sources. 15 missed cases (71.4 percent) were found in the MRDI, 6 (28.6 percent) were found in pathology reports, and 1 (4.8 percent) was found in cytology reports.

Table 1A. Number of Missed Cases by Casefinding Source—Summary Report

Total Missed Cases*	MRDI	Path	RT Log	Cyto	Autopsy	Op Log	Oth
21	15	6	0	1	0	0	0

Summary of missed cases by casefinding source: One source = 20 cases; two sources = 1 case; three sources = none, and four or more sources = none.

***** Some cases were missed in multiple sources.

MRDI = Medical Record Disease Indices
Path = Pathology Report
RT Log = Radiation Therapy Log
Cyto = Cytology Report
Op Log = Surgery/Operation Log
Oth = Other Sources

Table 1B shows the number of missed cases by primary site found in each source reviewed. Among the 21 missed cases, 9 (42.9 percent) were found in the high-caseload facilities, 8 (38.1 percent) were found in the medium-caseload facilities, and 4 (19.0 percent) were found in the low-caseload facilities.

Lung and bronchus and colon and rectum cancer cases were missed most often. There were 9 missed lung and bronchus cancer cases (42.9 percent), 7 missed colon and rectum cancer cases (33.3 percent), 3 missed prostate cancer cases (14.3 percent) and 2 missed female breast cancer cases (9.5 percent).

The high-caseload facilities missed the most number of cases. There was a total of 9 missed cases in this category. Of the 9 missed cases, 6 (66.7 percent) were found in the MRDI alone, and 3 (33.3 percent) were found in pathology reports alone. There were no missed cases in other casefinding sources for the high-caseload facilities. Lung cancer cases accounted for the highest number of missed cases among the high-caseload facilities with 5 missed cases (55.6 percent), followed by female breast cancer cases with 2 (22.2 percent) and colon and rectum and prostate cancer with 1 each (11.1 percent each).

Table 1B. Number of Missed Cases, by Primary Site, Casefinding Source, and Registry Status—Detailed Report

Primary Site	Total Missed Cases per Site	No. of Missed Cases per Source	Missed Cases by No. of Sources	MRDI	Path	RT Log	Cyto	Autopsy	OP Log	Oth
High/Registry (total missed cases = 9)										
Female Breast	2	1	1		X					
		1	1	X						
Colon & Rectum	1	1	1	X						
Lung	5	4	1	X						
		1	1		X					
Prostate	1	1	1		X					
Medium/Registry (total missed cases = 8)										
Colon & Rectum	4	2	1		X					
		2	1	X						
Lung	3	3	1	X						
Prostate	1	1	1	X						
Low/Registry (total missed cases = 4)										
Colon & Rectum	2	2	1	X						
Lung	1	1	2		X		X			
Prostate	1	1	1	X						

Registry = Hospital with in-house cancer registry
MRDI = Medical Record Disease Indices
Path = Pathology Report
RT Log = Radiation Therapy Log
Cyto = Cytology Report
Op Log = Surgery/Operation Log
Oth = Other Sources

In the medium-caseload facilities, 6 missed cases (75.0 percent) were found in the MRDI alone, and 2 (25.0 percent) were found in pathology reports alone. There were no missed cases in other casefinding sources for the medium-caseload facilities. Colon and rectum cancer cases accounted for the highest number of missed cases in the medium-caseload facilities with 4 missed cases (50.0 percent), lung and bronchus cancer had 3 missed cases (37.5 percent), and prostate cancer had 1 missed case (12.5 percent). There were no missed cancer cases for female breast cancer.

The lowest number of missed cases was found in the low-caseload facilities. Of the 4 missed cases, 2 were identified in the MRDI alone (50.0 percent), 1 case was identified in pathology alone (25.0 percent), and 1 case was identified in both pathology and cytology reports (25.0 percent). There were no missed cases found in any other casefinding sources. There were

2 missed cases (50.0 percent) among colon and rectum cancer cases; 1 case (25.0 percent) was missed in lung and bronchus, and 1 case (25.0 percent) was missed in prostate. There were no missed cases for female breast cancer.

Table 1C. Case Completeness, by Caseload Category and Registry Status

Caseload/Registry Status	Number of Facilities in the Sample	Number of Missed Cases	Percentage Case Completeness	95% Confidence Interval
High/Registry	4	9	97.6	±0.34
Medium/Registry	3	8	97.4	±0.32
Low/Registry	2	4	97.5	±0.44
Total	9	21	97.5	±0.65

The overall case ascertainment completeness for ACR was estimated at 97.5 percent (96.85 percent to 98.15 percent, with a 95 percent confidence interval of ±0.65 percent). The high-caseload facilities attained a case ascertainment completeness of 97.6 percent (97.26 percent to 97.94 percent, with a 95 percent confidence interval of ±0.34 percent). The medium-caseload facilities had a case ascertainment completeness of 97.4 percent (97.08 percent to 97.72 percent, with a 95 percent confidence interval of ±0.32 percent). The low-caseload facilities had a case ascertainment completeness of 97.5 percent (97.06 percent to 97.94 percent, with a 95 percent confidence interval of ±0.44 percent).

CDC's national standard for central cancer registry completeness states that "ninety-five percent of unduplicated, expected, malignant cases of reportable cancer occurring in State residents should be reported within 24 months of the close of each diagnosis year." **ACR's overall estimated case completeness of 97.5 percent surpasses the NPCR standard for the cancer sites audited. The Arizona Cancer Registry is to be commended for this outstanding result.**

DATA QUALITY

A total of 297 records were reabstracted for the data quality portion of the audit. As shown in table 2A, 132 reabstracted records were from the high-caseload facilities, 99 were from the medium-caseload facilities, and 66 were from the low-caseload facilities. Data discrepancies were identified in 162 records, representing 54.5 percent of all records reviewed. 69 records with discrepancies were from the high-caseload facilities, representing 42.6 percent of all records with data discrepancies, 56 (34.6 percent) were from the medium-caseload facilities, and 37 (22.8 percent) were from the low-caseload facilities. A total of 135 records (45.5 percent) was error free. 47.7 percent of the records reviewed in the high-caseload facilities were error free, followed by the low-caseload facilities with 43.9 percent, and the medium-caseload facilities with 43.4 percent.

Table 2A. Records With and Without Discrepancies, by Hospital Caseload Category and Registry Status

Registry Status	No. of Records Reabstracted	No. of Records With Discrepancy	No. of Records Without Discrepancy	Percentage of Records Error Free
High/Registry 4 Facilities	132	69	63	47.7
Medium/Registry 3 Facilities	99	56	43	43.4
Low/Registry 2 Facilities	66	37	29	43.9
Total	297	162	135	45.5

Table 2B describes data elements with and without discrepancies. A total of 13 critical data elements was examined in each record. Out of a total of 3,861 data elements, 254 data elements (6.6 percent) were found to have discrepancies.

- 103 discrepancies were in the high-caseload facilities (40.6 percent)
- 88 discrepancies were in the medium-caseload facilities (34.6 percent)
- 63 data discrepancies were in the low-caseload facilities (24.8 percent)

The resultant aggregate data accuracy rate was 93.4 percent.

Table 2B. Data Elements Reviewed With and Without Discrepancies, by Hospital Caseload Category and Registry Status

Registry Status	No. of Critical Data Elements Reviewed	No. of Critical Data Elements With Discrepancy	No. of Critical Data Elements Without Discrepancy	Percentage of Critical Data Elements Error Free
High/Registry 4 Facilities	1,716	103	1,613	94.0
Medium/Registry 3 Facilities	1,287	88	1,199	93.2
Low/Registry 2 Facilities	858	63	795	92.7
Total	3,861	254	3,607	93.4

The overall data accuracy rate for ACR was estimated at 93.4 percent (92.61 percent to 94.19 percent, with a 95 percent confidence interval of ±0.79 percent). The data accuracy was 94.0 percent (93.52 percent to 94.48 percent, with a 95 percent confidence interval of ±0.48 percent) for the high-caseload facilities, 93.2 percent (92.76 percent to 93.64 percent, with a 95 percent confidence interval of ±0.44 percent) for the medium-caseload facilities, and 92.7 percent (92.26 percent to 93.14 percent, with a 95 percent confidence interval of ±0.44 percent) for the low-caseload facilities (table 2C).

Table 2C. Data Quality, by Caseload Category and Registry Status

Caseload/Registry	Percentage of Data Accuracy	95% Confidence Interval
High/Registry 4 Facilities	94.0	±0.48
Medium/Registry 3 Facilities	93.2	±0.44
Low/Registry 2 Facilities	92.7	±0.44
Total	93.4	±0.79

ACR's overall data accuracy rate for the four cancer sites audited was 93.4 percent.

Tables 3A through 3D and 4A through 4D in the appendix describe discrepancies in data quality in the 13 critical data elements examined. Tables 3A through 3D further break down the discrepancies into major and minor discrepancies. Tables 4A through 4D represent a breakdown of the discrepancies without distinguishing between major and minor discrepancies. Figure 1 in the appendix presents the percentage of errors by primary site for ACR, and figure 2 compares the percentage of errors in each variable among primary sites for ACR and NPCR 1993–1999. This report focuses on tables 3 and 3A through 3D.

Major discrepancies are the errors in the critical data elements that affect the incidence or survival rate. Minor discrepancies are the errors in the critical data elements that may not directly affect those rates.

Errors in *Stage at Diagnosis, Race, State of Residence, Primary Site, Histology, Sequence Number, Laterality, Gender,* and *Behavior* are considered major. Errors in *Grade* and *Subsite* are considered minor. Discrepancies in *Date of Diagnosis (mm/dd<=30 days)* and *Date of Birth (mm/dd<=30 days)* are considered minor, while discrepancies in *Date of Diagnosis (yy)* or *Date of Diagnosis (mm/dd>30 days)* and *Date of Birth (yy)* or *Date of Birth (mm/dd>30 days)* are considered major.

Lung and bronchus cancer cases had the highest number of discrepancies with 97, representing 38.2 percent of all data discrepancies. Female breast cancer cases had 71 discrepancies (28.0 percent), colon and rectum cancer cases had 58 (22.8 percent), and prostate cancer cases had 28 (11.0 percent). Major and minor discrepancies for ACR are represented in tables 3A through 3D, along with the NPCR and SEER program averages. There were 254 data discrepancies. 136 discrepancies were considered major (53.5 percent), and 118 discrepancies were considered minor (46.5 percent).

Among the 136 **major discrepancies,** lung and bronchus cancer cases had the most major errors with 49 (36.0 percent), colon and rectum cancer cases had 37 (27.2 percent), female breast cancer cases had 36 (26.5 percent), and prostate cancer cases had 14 (10.3 percent). *Stage at Diagnosis* was the data element that had the highest number of discrepancies with 57 (41.9 percent). This was followed by 27 discrepancies (19.9 percent) in *Histology,* 13 (9.6 percent) in *Laterality,* and 10 each (7.4 percent each) in *Date of Diagnosis (yy or mm/dd>30*

days) and *Race*. 7 discrepancies (5.1 percent) were identified in *Primary Site.* 5 discrepancies (3.7 percent) were identified in *Sequence Number,* 4 discrepancies (2.9 percent) were identified in *Behavior,* 2 discrepancies (1.5 percent) were identified in *State of Residence,* and 1 discrepancy (0.7 percent) was identified in *Date of Birth (yy or mm/dd>30 days). Gender* had no discrepancies (tables 3 and 3A through 3D).

Stage at Diagnosis had a total of 57 discrepancies, of which 24 (42.1 percent) were identified in the medium-caseload facilities. 19 discrepancies (33.3 percent) were identified in the high-caseload facilities, and 14 (24.6 percent) were found in the low-caseload facilities.

Histology had a total of 27 discrepancies, of which 10 (37.0 percent) were identified in the high-caseload facilities, 9 (33.3 percent) were identified in the medium-caseload facilities, and 8 (29.6 percent) were identified in the low-caseload facilities.

There were 13 discrepancies identified in *Laterality,* 7 of which (53.8 percent) were found in the medium-caseload facilities. 5 discrepancies (38.5 percent) were found in the high-caseload facilities and 1 (7.7 percent) in the low-caseload facilities.

Date of Diagnosis (yy or mm/dd>30 days) had 6 discrepancies in the high-caseload facilities, which represents 60.0 percent of the 10 discrepancies in this data element. The low-caseload facilities had 2 discrepancies (20.0 percent), and the high-caseload facilities had 2 discrepancies (20.0 percent).

Race had 10 discrepancies, of which 5 (50.0 percent) were found in the medium-caseload facilities, 4 (40.0 percent) in the high-caseload facilities, and 1 (10.0 percent) in the low-caseload facilities.

Primary Site had a total of 7 discrepancies, of which 3 (42.9 percent) were identified in the high-caseload facilities and 2 each (28.6 percent each) in the medium- and low-caseload facilities.

Sequence Number had 5 discrepancies, of which 3 (60.0 percent) were found in the medium-caseload facilities, 1 (20.0 percent) was found in the low-caseload facilities, and 1 (20.0 percent) was found in the high-caseload facilities.

Behavior had 4 discrepancies, of which 2 (50.0 percent) were found in the high-caseload facilities and 2 (50.0 percent) were found in the low-caseload facilities. There were no discrepancies in the medium-caseload facilities for this data element.

State of Residence had 2 discrepancies, of which 1 (50.0 percent) was found in the medium-caseload facilities and 1 (50.0 percent) was found in the low-caseload facilities. There were no discrepancies in the high-caseload facilities for this data element.

There was 1 discrepancy (100.0 percent) in *Date of Birth (yy or mm/dd>30 days),* which occurred in the low-caseload facilities. There were no discrepancies identified in the high- or medium-caseload facilities for this data element.

There were no discrepancies identified in *Gender.*

Among the 118 **minor discrepancies,** *Subsite* had the highest number with 43 (36.4 percent). *Grade* had 40 discrepancies (33.9 percent), and *Date of Diagnosis (mm/dd<=30 days)* had 32 discrepancies (27.1 percent). There were 3 discrepancies (2.5 percent) in *Date of Birth (mm/dd<=30 days).*

Subsite had 43 discrepancies, of which 19 (44.2 percent) were identified in the high-caseload facilities, 15 (34.9 percent) were identified in the low-caseload facilities, and 9 (20.9 percent) were identified in the medium-caseload facilities.

40 discrepancies were identified in *Grade*, with 19 (47.5 percent) in the high-caseload facilities, 14 (35.0 percent) in the medium-caseload facilities, and 7 (17.5 percent) in the low-caseload facilities.

Among the 32 discrepancies in *Date of Diagnosis (mm/dd<=30 days),* 15 (46.9 percent) were identified in the high-caseload facilities. There were 11 discrepancies (34.4 percent) in the medium-caseload facilities, and the low-caseload facilities had 6 discrepancies (18.8 percent).

There were 3 discrepancies in *Date of Birth (mm/dd<=30 days)*, 2 (66.7 percent) of which were found in the low-caseload facilities and 1 (33.3 percent) in the medium-caseload facilities. There were no discrepancies for this data element in the high-caseload facilities.

FEMALE BREAST CANCER

Tables 3A and 4A represent the data discrepancies in the female breast cancer cases. Female breast cancer cases had the second highest number of discrepancies among the four cancer sites audited with 71 (28.0 percent). The medium-caseload facilities had 31 discrepancies (43.7 percent), the high-caseload facilities had 25 discrepancies (35.2 percent), and the low-caseload facilities had 15 discrepancies (21.1 percent). 36 of the discrepancies (50.7 percent) were considered major; 35 (49.3 percent) were considered minor (table 3A).

Among the 36 **major discrepancies,** the highest number was found in *Stage at Diagnosis* with 15 discrepancies (41.7 percent). *Histology* had 9 discrepancies (25.0 percent), *Race* had 4 discrepancies (11.1 percent), *Date of Diagnosis (yy or mm/dd>30 days)* had 3 discrepancies (8.3 percent), and *Behavior* had 2 discrepancies (5.6 percent). *Date of Birth (yy or mm/dd>30 days), State at Residence, Sequence Number* and had 1 discrepancy each (2.8 percent each). *Primary Site, Laterality,* and *Gender* had no discrepancies in this site.

Stage at Diagnosis had 11 discrepancies (73.3 percent) in the medium-caseload facilities. There were 2 discrepancies (13.3 percent) in the high-caseload facilities and 2 discrepancies (13.3 percent) in the low-caseload facilities.

- 9 discrepancies (60.0 percent) were recoded to a less extensive stage of disease. 7 of these 9 cases were recoded from distant stage of disease to regional lymph node involvement only. 1 case was recoded from distant stage of disease to localized, and 1 case was recoded from regional by direct extension to localized. In each case the information necessary to assign stage was available in the medical record and no documentation could be found that supported the distant stage.

- 3 discrepancies (20.0 percent) were recoded from a less extensive stage of disease to a more extensive stage of disease. 2 of the cases were recoded from in situ to localized based on pathology report documentation of invasion and microinvasion, and 1 case was recoded from localized to regional by lymph node based on pathology documentation of lymph node involvement.

- 2 discrepancies (13.3 percent) were recoded an unknown stage of disease to a known stage of disease. 1 of the cases was recoded from unknown to distant based on a positive bone scan, and 1 was recoded from unknown to localized based on clinically clear lymph nodes and grossly clear margins.

- 1 discrepancy (6.7 percent) was recoded from localized to an unknown stage of disease when the primary was found to be recurrence from 1989.

Histology had 4 discrepancies each (44.4 percent each) in the high- and low-caseload facilities and 1 discrepancy (11.1 percent) in the medium-caseload facilities.

- 6 of these discrepancies (66.7 percent) were recoded from a lower histology to a higher histology. 3 cases were recoded from ductal carcinoma (8500) to comedocarcinoma (8501), 2 were recoded from ductal (8500) or lobular (8520) carcinoma to lobular and ductal carcinoma combination code (8522), and 1 case was recoded from carcinoma in situ (8010) to ductal carcinoma in situ (8500). All of these changes were made based on pathology report documentation and application of mixed histology rules in place for year 2000 cases.

- 3 of these discrepancies (33.3 percent) were recoded from a higher histology to a lower histology. 1 case was recoded from intraductal and lobular carcinoma in situ (8522) to lobular carcinoma in situ (8520) when no evidence for ductal carcinoma could be found, 1 was recoded from intraductal and lobular carcinoma in situ (8522) to ductal carcinoma in situ (8500) when no evidence for lobular carcinoma could be found, and 1 case was recoded from intraductal papillary carcinoma (8503) to papillary carcinoma in situ (8050) when outside pathology reports did not support any reference to ductal.

Race had 3 discrepancies (75.0 percent) in the medium-caseload facilities and 1 discrepancy (25.0 percent) in the high-caseload facilities. There were no discrepancies in this data element in the low-caseload facilities.

- 2 discrepancies were recoded from white (01) to either Japanese (05) or Southeast Asian (96) based on medical record documentation.

- 1 discrepancy was recoded from white (01) to an unknown race (99) when the medical record had no documentation of race.

- 1 discrepancy was recoded from white (01) to other (98) based on the patient's self identification found in the chart.

Date of Diagnosis (yy or mm/dd>30 days) had 2 discrepancies (66.7 percent) in the high-caseload facilities and 1 (33.3 percent) in the medium-caseload facilities. There were no discrepancies for this data element in the low-caseload facilities.

- 2 discrepancies were recoded to an earlier diagnosis date. 1 case was recoded to greater than 10 years earlier when it was documented to be a recurrence, and 1 case was recoded to more than a month earlier when the mammogram showed diagnostic language.

- 1 discrepancy was recoded from an unspecified date in the year 2000 to the admission date of the procedure.

Behavior had 1 discrepancy (50.0 percent) in the high-caseload facilities and 1 discrepancy (50.0 percent) in the low-caseload facilities. There were no discrepancies in the medium-caseload facilities for this data element. These 2 cases were recoded from in situ to invasive based on pathology report documentation of invasion and microinvasion.

Sequence Number had 1 discrepancy in the medium-caseload facilities. There were no discrepancies in this data element in the high- or low-caseload facilities. The single discrepancy was recoded from sequence 02 to sequence 00 when it was documented to be a recurrence of an earlier breast primary.

Date of Birth (yy or mm/dd>30 days) had 1 discrepancy in the low-caseload facilities. There were no discrepancies in the high or medium-caseload facilities. The 1 discrepancy in *Date of Birth (yy or mm/dd>30 days)* was recoded to a later date based on face sheet documentation.

State of Residence had 1 discrepancy, which was identified in the medium-caseload facilities. There were no discrepancies in the high- or low-caseload facilities for this data element. This case was recoded from an Arizona resident to unknown resident when the diagnosis date was documented to be greater than 10 years prior to reporting.

There were no discrepancies in *Primary Site, Laterality,* or *Gender.*

Among the 35 **minor discrepancies,** *Subsite* had the highest number with 17 (48.6 percent), followed by *Grade* with 12 (34.3 percent), *Date of Diagnosis (mm/dd<=30 days)* with 5 (14.3 percent), and *Date of Birth (mm/dd<=30 days)* with 1 (2.9 percent).

Subsite had 8 discrepancies (47.1 percent) in the high-caseload facilities, 5 (29.4 percent) in the medium-caseload facilities, and 4 (23.5 percent) in the low-caseload facilities.

- 5 cases were recoded from a specific subsite of the breast (C50.X) to an overlapping lesion (C508) based on biopsy information, operative reports, and physical examinations.

- 4 cases were recoded from a nonspecific subsite code (C50.9) to a specific subsite (C50.X) when biopsy reports, needle localization reports, and history and physical reports documented a specific subsite.

- 4 cases were recoded from a specific subsite of the breast (C50.X) to a different subsite of the breast (C50.X) when radiology reports, pathology reports, history and physical reports, and surgical evaluations documented a different subsite.

- 3 cases were recoded from an overlapping lesion (C50.8) to a specific subsite (C50.X) when pathology, history and physical reports, and operative reports showed a specific subsite designated.

- 1 case was recoded from central breast (C50.1) to unspecified breast (C50.9) when no documentation could be found to support the central code.

A focused review of the breast clock diagram and development of a policy on the hierarchy of sources to be used for subsite determination, similar to the policy in place for summary staging (the SEER Summary Staging Guide and SEER Program Manual), would help to reduce these errors.

Grade had 12 discrepancies, of which 5 (41.7 percent) were identified in the medium-caseload facilities, 4 (33.3 percent) were identified in the high-caseload facilities, and 3 (25.0 percent) were identified in the low-caseload facilities.

- 8 discrepancies were recoded from a lower grade to a higher grade. 4 cases involved applying Scarff Bloom-Richardson (SBR) grading scheme conversions to cases where no SBR information was documented in the pathology report. 3 cases involved pathology documentation of more than one grade, and the higher grade was selected during the audit recode; 1 case was recoded from well differentiated to moderately differentiated when no documentation of well differentiated could be found.

- 3 discrepancies were recoded from an unknown grade to a known grade when pathology report documentation provided a grade.

- 1 discrepancy was recoded from poorly differentiated to unknown grade when documentation showed the grade was taken from the biopsy of a metastatic site.

Date of Diagnosis (mm/dd<=30 days) had 3 discrepancies (60.0 percent) in the high-caseload facilities and 2 discrepancies (40.0 percent) in the medium-caseload facilities. There were no discrepancies in the low-caseload facilities for this data element.

- 3 discrepancies were recoded from an earlier diagnosis date to a later diagnosis date based on positive biopsy reports and nondiagnostic language in the mammograms.

- 2 discrepancies were recoded from a later diagnosis date to an earlier diagnosis date based on documentation of clinical diagnosis of malignancy.

Date of Birth (mm/dd<=30 days) had 1 discrepancy, which was found in the medium-caseload facilities. The high- and low-caseload facilities did not have any discrepancies for this data element. The discrepancy was recoded to a later date based on face sheet documentation.

COLON AND RECTUM CANCER

Tables 3B and 4B in the appendix represent discrepancies in colon and rectum cancer cases. Colon and rectum cancer cases had the third highest number of discrepancies among the four cancer sites audited. A total of 58 discrepancies was identified for this site. 27 (46.6 percent) of these discrepancies were identified in the high-caseload facilities, 16 (27.6 percent) in the medium-caseload facilities, and 15 (25.9 percent) in the low-caseload facilities. 37 of the 58 discrepancies (63.8 percent) classified as major and 21 (36.2 percent) classified as minor (table 3B).

Among the 37 **major discrepancies,** *Stage at Diagnosis* had 18 discrepancies, representing 48.6 percent of major discrepancies. *Histology* had 7 discrepancies (18.9 percent), *Primary Site* had 5 discrepancies (13.5 percent), *Laterality* and *Behavior* had 2 discrepancies each (5.4 percent each), and *Race, State of Residence,* and *Sequence Number* had 1 discrepancy each (2.7 percent each). There were no discrepancies in *Date of Diagnosis (yy or mm/dd>30 days), Date of Birth (yy or mm/dd>30 days),* or *Gender.*

Stage at Diagnosis had 7 discrepancies (38.9 percent) in the high-caseload facilities, 6 discrepancies (33.3 percent) in the low-caseload facilities, and 5 discrepancies (27.8 percent) in the medium-caseload facilities.

- 9 discrepancies were recoded from a more extensive stage of disease to a less extensive stage of disease. 6 of these cases were recoded from regional by direct extension to localized when pathology reports showed no invasion into the serosa, 1 was recoded from distant to regional by lymph nodes only when liver metastasis was established after the 2-month staging timeframe, 1 was recoded from regional by direct extension and lymph nodes to regional by lymph nodes only when pathology reports documented no invasion of serosa, and 1 case was recoded from local to in situ when the pathology report documented no invasive carcinoma.

- 4 discrepancies were recoded from an unknown stage to a known stage. 2 of these cases were recoded from unknown to localized based on resection pathology reports and staging scans available in the chart, 1 case was recoded from unknown to regional by direct extension based on colonoscopy documentation that the tumor was fixed, and 1 case was recoded to distant based on documented liver metastasis.

- 3 discrepancies were recoded from a less extensive stage of disease to a more extensive stage of disease. 2 cases were recoded from localized to regional by direct extension based on pathology report documentation of serosal involvement or perforation, and 1 was recoded from in situ to localized based on resection pathology report documentation of invasion.

- 2 discrepancies were recoded from a known stage of disease to an unknown stage of disease when no documentation could be found to support the codes available. Both cases involved tumors without resections available.

Histology had 3 discrepancies each (42.9 percent each) in the high- and medium-caseload facilities and 1 discrepancy (14.3 percent) in the low-caseload facilities.

- 4 discrepancies were recoded from a lower histology to a higher histology. 2 of these cases were recoded from adenocarcinoma (8140) to either cancer in a villous adenoma (8261) or adenocarcinoma in polyp (8210) based on pathology report documentation, 1 was recoded from adenocarcinoma in villous adenoma (8261) to adenocarcinoma in tubulovillous adenoma (8263) based on pathology report documentation, and 1 case was recoded from adenocarcinoma in tubulovillous adenoma (8263) to mucinous adenocarcinoma (8480) based on pathology report documentation.

- 3 discrepancies were recoded from a higher histology to a lower histology. Of these 3 cases, 1 was recoded from mucinous adenocarcinoma (8480) to adenocarcinoma (8140) when pathology report documentation showed only focal areas of colloid carcinoma, 1 case was recoded from adenocarcinoma (8140) to cloacogenic carcinoma (8124) when pathology report documentation could not be found for the code submitted, and 1 case was recoded from adenocarcinoid tumor (8245) to carcinoid NOS not of appendix (8240) based on pathology report documentation.

Primary Site had 2 discrepancies each (40.0 percent each) in the high- and medium-caseload facilities and 1 discrepancy (20.0 percent) in the low-caseload facilities.

- 2 cases were recoded to rectosigmoid junction (C19.9) from either rectum (C20.9) or overlapping colon (C18.8) based on colonoscopy and operative reports that placed the origin of the tumor at 12 cm.

- 1 case was recoded from a cecal primary (C18.0) to an unknown primary (C80.9) based on pathology and discharge summary documentation of unknown origin.

- 1 case was recoded from rectosigmoid junction (C19.9) to rectum (C20.9) based on physician documentation of site and supporting documentation that the mass was palpable on digital rectal examination.

- 1 case was recoded from rectum (C20.9) to anal canal (C21.1) based on pathology report documentation of site, histology associated with anal primary, and physician documentation of anal primary.

Laterality had 2 discrepancies, both of which (100.0 percent) occurred in the medium-caseload facilities. There were no discrepancies in the high- and low-caseload facilities for this data element. The 2 discrepancies were recoded from right (01) to not a paired organ (0) based on the site (rectum and ascending colon).

Behavior had 2 discrepancies, of which 1 (50.0 percent) was found in the high-caseload facilities and 1 (50.0 percent) was found in the low-caseload facilities. There were no discrepancies identified in the medium-caseload facilities for this data element.

- 1 case was recoded from in situ (2) to invasive (3) based on documentation in the pathology report.

- 1 case was recoded from invasive (3) to in situ (2) based on documentation in the pathology report.

Race had 1 discrepancy (100.0 percent), which was identified in the high-caseload facilities. This case was recoded from white (01) to unknown (99) when no documentation of race could be found in the record.

State of Residence had 1 discrepancy found in the low-caseload facilities. This case was recoded from Arizona (AZ) to California (CA) when the chart showed documentation that the patient came to Arizona recently to visit his daughter and returned to California to have surgery.

Sequence Number had 1 discrepancy (100.0 percent) found in the medium-caseload facilities. This case was recoded from 00 to 01 when a sigmoid primary was noted on the pathology report in addition to the rectal primary reported.

There were no discrepancies found in *Date of Birth (yy or mm/dd>30 days), Date of Diagnosis (yy or mm/dd>30 days)* or *Gender.*

Among the 21 **minor discrepancies** in the colon and rectum cancer cases, *Subsite* had 10 discrepancies (47.6 percent), *Grade* had 6 (28.6 percent), and *Date of Diagnosis (mm/dd<=30 days)* had 5 (23.8 percent). *Date of Birth (mm/dd<=30 days)* had no discrepancies.

Subsite had 6 discrepancies (60.0 percent) identified in the high-caseload facilities and 4 discrepancies (40.0 percent) found in the low-caseload facilities. There were no discrepancies identified in the medium-caseload facilities for this data element.

- 3 discrepancies were recoded from the ascending colon (C18.2) to cecum (C18.0) based on pathology, operative, and colonoscopy report documentation.

- 3 discrepancies were recoded to transverse colon (C18.4) from either the flexures (C18.X) or descending colon (C18.6) based on operative report documentation.

- 3 discrepancies were recoded from a nonspecific code (C18.9) or (C18.8) to a specific code based on colonoscopy and operative reports.

- 1 discrepancy was recoded from cecum (C18.0) to ascending colon (C18.2) when colonoscopy first reported cecum but operative report stated in the area of the hepatic flexure and the discharge summary stated right colon.

Grade had 5 discrepancies (83.3 percent) in the high-caseload facilities and 1 (16.7 percent) in the medium-caseload facilities. There were no discrepancies in the low-caseload facilities for this data element.

- 3 discrepancies were recoded from an unknown grade (9) to a known grade based on pathology report and physician dictation of grade available in the medical record.

- 2 discrepancies were recoded from a lower grade to a higher grade based on pathology report documentation available in the medical record.

- 1 discrepancy was recoded from a known grade (4) to an unknown grade (9) when the biopsy was found to be of a metastatic site.

Date of Diagnosis (mm/dd<=30 days) had 2 discrepancies (40.0 percent) in the high-caseload facilities, 2 discrepancies (40.0 percent) in the medium-caseload facilities, and 1 (20.0 percent) in the low-caseload facilities.

- 4 discrepancies were recoded to an earlier date of diagnosis based on diagnostic language found in physician's dictation, notes, and colonoscopy reports.

- 1 discrepancy was recoded to a later date based on clinical diagnosis of cancer on endoscopy and no supporting documentation for the earlier date.

There were no discrepancies in *Date of Birth (mm/dd<=30 days).*

LUNG AND BRONCHUS CANCER

Tables 3C and 4C in the appendix show a total of 97 discrepancies among the lung and bronchus cancer cases, making it the site with the most data discrepancies. 33 discrepancies (34.0 percent) were identified in the high-caseload facilities, 32 (33.0 percent) were found in the medium-caseload facilities, and 32 (33.0 percent) were found in the low-caseload facilities. Of those, 49 (50.5 percent) were considered major discrepancies, and 48 (49.5 percent) were considered minor discrepancies (table 3C).

Among the 49 **major discrepancies,** the data element that had the highest number of discrepancies was *Stage at Diagnosis* with 20 discrepancies (40.8 percent), followed by *Histology* with 10 discrepancies (20.4 percent) and *Laterality* with 9 discrepancies (18.4 percent). *Date of Diagnosis (yy or mm/dd>30 days)* and *Sequence Number* had 3 discrepancies each (6.1 percent each), and *Primary Site and Race* had 2 discrepancies each (4.1 percent each). *Date of Birth (yy or mm/dd>30 days), State of Residence, Gender,* and *Behavior* had no discrepancies in this site.

Stage at Diagnosis had 8 discrepancies (40.0 percent) in the high-caseload facilities, 6 discrepancies (30.0 percent) in the medium-caseload facilities, and 6 (30.0 percent) in the low-caseload facilities.

- 12 discrepancies were recoded from an unknown stage to a known stage. 6 of these cases were recoded to localized extent of disease based on computerized axial tomography scans, chest x-rays, and physician staging, and 4 were recoded to distant disease based on physician's documentation of bone metastasis, bronchoscopic evidence of contralateral tumor spread, and computerized axial tomography evidence of bilateral pleural effusions. 1 was recoded to regional by direct extension based on documentation of laryngeal nerve involvement, and 1 was recoded to regional extension by lymph nodes only based on computerized axial tomography evidence of lymph node involvement.

- 5 discrepancies were recoded from a more extensive stage of disease to a less extensive stage. 3 of these cases were recoded from regional by direct extension and regional lymph

nodes to regional by lymph nodes only based on pathology reports and computerized axial tomography scans that showed no evidence of direct invasion. 1 case was recoded from regional by direct extension to localized when the pathology report confirmed no involvement of pleura. 1 case was recoded from distant to regional by direct extension and regional lymph nodes when no definitive evidence of distant metastasis could be found.

- 3 discrepancies were recoded from a less extensive extent of disease to a more extensive extent of disease. 2 cases were recoded to distant based on cytology reports and computerized axial tomography scans showing malignant pleural effusions, and 1 was recoded from localized to regional by lymph nodes only based on computerized axial tomography scans of the chest.

Histology had 4 discrepancies (40.0 percent) in the medium-caseload facilities and 3 discrepancies each (30.0 percent each) in the high- and low-caseload facilities.

- 8 discrepancies were recoded from a lower histology to a higher histology. Of these 8 cases, no clear pattern of discrepancies emerged, although a review of diagnostic language and careful attention to pathology and cytology reports would have reduced some of these discrepancies. 1 case was recoded from squamous cell carcinoma large non-keratinizing (8072) to adenocarcinoma (8140) when both histologies were documented in the pathology report; 1 case was recoded from adenocarcinoma (8140) to bronchiolar carcinoma (8250) when the autopsy report showed that the patient had two separate lung primaries and the histologies had been switched; 1 case was recoded from carcinoma, NOS (8010) to neuroendocrine carcinoma (8246) when the pathology report showed carcinoma with neuroendocrine features; 1 case was recoded from large cell carcinoma (8012) to adenocarcinoma (8140) when the pathology report showed results favoring adenocarcinoma; 1 case was recoded from malignancy, NOS (8000) to adenocarcinoma (8140) based on cytology report documentation of adenocarcinoma; 1 case was recoded from carcinoma, NOS (8010) to squamous cell carcinoma (8070) based on cytology documentation of probable squamous cell carcinoma; 1 case was recoded from large cell carcinoma (8012) to neuroendocrine carcinoma (8246) based on pathology documentation of neuroendocrine carcinoma with large cell component; and 1 case was recoded from small cell carcinoma (8041) to neuroendocrine carcinoma (8246) based on cytology documentation and no evidence for small cell found in pathology.

- 2 discrepancies were recoded from a higher histology to a lower histology. 1 case was recoded from adenosquamous carcinoma (8560) to carcinoma, NOS (8010) when pathology reports documented nonsmall cell carcinoma with rare areas suggestive of adenocarcinoma and squamous cell carcinoma, and 1 case was recoded from neuroendocrine carcinoma (8246) to carcinoid NOS not of appendix (8240) when pathology reports showed only documentation of a neuroendocrine neoplasm in conjunction with a carcinoid tumor.

Laterality had 5 discrepancies (55.6 percent) in the high-caseload facilities, 3 (33.3 percent) in the medium-caseload facilities, and 1 (11.1 percent) in the low-caseload facilities.

- 3 discrepancies were recoded from a known laterality to an unknown laterality when no documentation could be found in the medical record.

- 2 discrepancies were recoded from an unknown laterality to a known laterality when documentation was found on computerized axial tomography scans, biopsy, and history and physical reports.

- 1 discrepancy was recoded from a left-sided laterality (2) to both simultaneously (4) based on computerized axial tomography scans that documented bilateral masses suspicious for carcinoma without documentation of metastasis from one side to another.

- 1 discrepancy was recoded from both simultaneous (4) to left (2) based on bronchoscopic evidence of a left-sided malignancy and computerized axial tomography scans showing bilateral lymph node masses with only a left-sided parenchymal mass.

- 1 discrepancy was recoded from right laterality (1) to not applicable (0) when the primary site was documented to be trachea with lung metastasis.

- 1 discrepancy was recoded from an unknown laterality (9) to right (1) when the primary was documented to be the right breast instead of lung.

Date of Diagnosis (yy or mm/dd>30 days) had 1 discrepancy each (33.3 percent each) in the high-, medium-, and low-caseload facilities.

- 1 discrepancy was recoded to an earlier date based on diagnostic language in a computerized axial tomography scan found in the radiation oncology chart.

- 1 discrepancy was recoded to 99/99/00 from 2/2/00 when the lung primary was discovered to be metastasis from a laryngeal primary.

- 1 discrepancy was recoded to 1 month earlier when diagnostic language appeared in an emergency room record.

Sequence Number had 1 discrepancy each (33.3 percent each) in the high-, medium-, and low-caseload facilities.

- 1 discrepancy was recoded from 00 to 01 when two other primaries were discovered in the medical record.

- 1 discrepancy was recoded from 02 to 00 when the lung primary was found to be metastasis from a tracheal primary.

- 1 discrepancy was recoded from 01 to 00 when the lung malignancy was shown to be metastasis from a breast primary.

Primary Site had 1 discrepancy (50.0 percent) in the high-caseload facilities and 1 discrepancy (50.0 percent) in the low-caseload facilities. There were no discrepancies in the medium-caseload facilities for this data element.

- 1 discrepancy was recoded from a left upper lobe lung primary (C34.1) to a tracheal primary after review of merged record.

- 1 discrepancy was recoded from an overlapping lung primary (C348) to the LOQ of the breast (C505).

Race had 1 discrepancy each (50.0 percent each) in the medium- and low-caseload facilities. There were no discrepancies in the high-caseload facilities for this data element.

- 1 discrepancy was recoded from white (01) to unknown (99) when conflicting documentation of race was documented in the medical record.

- 1 discrepancy was recoded from an unknown race (99) to white (01) based on face sheet documentation.

There were no discrepancies found for *Date of Birth (yy or mm/dd>30 days), State of Residence, Gender,* and *Behavior*.

Among the 48 **minor discrepancies,** *Subsite* and *Date of Diagnosis (mm/dd<=30 days)* had 16 discrepancies each (33.3 percent each), *Grade* had 14 discrepancies (29.2 percent), and *Date of Birth (mm/dd<=30 days)* had 2 discrepancies (4.2 percent).

Subsite had 7 discrepancies (43.8 percent) in the low-caseload facilities, 5 (31.3 percent) in the high-caseload facilities, and 4 (25.0 percent) in the medium-caseload facilities.

- 6 of these discrepancies were recoded from an overlapping lesion (C34.8) to a more specific subsite. Of these 6, 4 were recoded from an overlapping lesion (C34.8) to upper lobe (C34.1), and 2 were recoded from an overlapping lesion (C34.8) to lower lobe (C34.3). All changes were made based on documentation in bronchoscopies, operative reports, and scans.

- 4 discrepancies were recoded from a specific subsite (C34.X) to a different subsite (C34.X) based on bronchoscopic and operative report documentation.

- 4 discrepancies were recoded from an unknown subsite (C34.9) to a specific subsite (C34.X) based on bronchoscopic reports, computerized axial tomography reports, and physician documentation.

- 2 discrepancies were recoded from a specific subsite (C34.X) to an unknown subsite (C34.9) when no documentation could be found to substantiate the subsite code chosen.

Date of Diagnosis (mm/dd<=30 days) had 6 discrepancies (37.5 percent) in the medium-caseload facilities and 5 discrepancies each (31.3 percent each) in the high- and low-caseload facilities.

- 13 of these discrepancies were recoded to an earlier diagnosis date than submitted. All of these cases were recoded based on diagnostic language found in magnetic resonance imaging scans, computerized axial tomography scans, physician's dictation, and handwritten progress notes.

- 3 discrepancies were recoded to a later diagnosis date than submitted. All of these cases were recoded when scans used for the diagnosis date did not contain the necessary diagnostic language.

Grade had 6 discrepancies (42.9 percent) in the medium-caseload facilities and 4 discrepancies each (28.6 percent each) in the high- and low-caseload facilities.

- 11 of these discrepancies were recoded from a known grade to an unknown grade. All of these cases were recoded when the grade was found to be taken from a metastatic site.

- 2 discrepancies were recoded from a lower grade (3) to a higher grade (4) based pathology documentation of undifferentiated.

- 1 discrepancy was recoded from undifferentiated (4) to poorly differentiated (3) when only documentation for grade 3 could be found in the medical record.

Date of Birth (mm/dd<=30 days) had 2 discrepancies, both found in the low-caseload facilities. There were no discrepancies for the high- and medium-caseload facilities in this data element. Both discrepancies were recoded to a later date based on face sheet documentation.

PROSTATE CANCER

Tables 3D and 4D in the appendix represent discrepancies among prostate cancer cases. The lowest number of data discrepancies, a total of 28, was found in the prostate cancer cases. 18 discrepancies (64.3 percent) were identified in the high-caseload facilities, 9 discrepancies (32.1 percent) were identified in the medium-caseload facilities, and 1 (3.6 percent) was identified in the low-caseload facilities. 14 discrepancies (50.0 percent) were categorized as major, and 14 (50.0 percent) were categorized as minor (table 3D).

Among the 14 **major discrepancies,** *Stage at Diagnosis* and *Date of Diagnosis (yy or mm/dd>30 days)* had 4 discrepancies each (28.6 percent each), followed by *Race* with 3 discrepancies (21.4 percent), *Laterality* with 2 discrepancies (14.3 percent), and *Histology* with 1 (7.1 percent). *Primary Site, Date of Birth (yy or mm/dd>30 days), State of Residence, Sequence Number, Gender,* and *Behavior* had no discrepancies.

Date of Diagnosis (yy or mm/dd>30 days) had 3 discrepancies (75.0 percent) in the high-caseload facilities and 1 (25.0 percent) in the low-caseload facilities. There were no discrepancies in the medium-caseload facilities for this data element.

- 2 discrepancies were recoded from 99/99/2000 to a more specific date after review of the medical record.

- 1 discrepancy was recoded to an earlier date when a bone scan showed an earlier diagnosis.

- 1 discrepancy was recoded to a later date based on the admission date.

Stage at Diagnosis had 2 discrepancies each (50.0 percent each) in the high- and medium-caseload facilities. There were no discrepancies in the low-caseload facilities for this data element.

- 2 discrepancies were recoded from localized extent of disease to unknown extent of disease when no work up of any kind was found. Both cases involved Transurethal surgery (TRUS) biopsies only without physical examinations, scans, or histories available.

- 2 discrepancies were recoded from unknown extent of disease to a known extent of disease. 1 of these cases was recoded from unknown (9) to localized based on the resection pathology report, and 1 was recoded from unknown (9) to regional by direct extension when the resection pathology report showed extension through the prostatic capsule.

Race had 2 discrepancies (66.7 percent) in the high-caseload facilities and 1 discrepancy (33.3 percent) in the medium-caseload facilities. There were no discrepancies in the low-caseload facilities for this data element.

- 2 discrepancies were recoded from white (01) to unknown race (99) when no documentation of race was found in the record.

- 1 discrepancy was recoded from white (01) to Native American (3) based on face sheet documentation.

Laterality had 2 discrepancies, both of which (100.0 percent) were found in the medium-caseload facilities. There were no discrepancies found in the high- and low-caseload facilities for this data element. Both cases were recoded from a paired laterality to not applicable (0) since prostate is not a paired organ.

Histology had 1 discrepancy (100.0 percent), which was found in the medium-caseload facilities. This case was recoded from carcinoma, NOS (8010) to malignancy, NOS (8000) when no pathology was found on the chart and the term cancer was the only documentation found.

Among the 14 **minor discrepancies,** 8 (57.1 percent) occurred in *Grade*. 6 (42.9 percent) occurred in *Date of Diagnosis (mm/dd<=30 days)*. *Date of Birth (mm/dd<=30 days)* and *Subsite* did not have any discrepancies.

Grade had 6 discrepancies (75.0 percent) in the high-caseload facilities and 2 discrepancies (25.0 percent) in the medium-caseload facilities. There were no discrepancies in the low-caseload facilities for this data element.

- 3 discrepancies were recoded from a lower grade to a higher grade. All of these cases were recoded from moderately differentiated (2) to poorly differentiated (3) when multiple grading systems were used and the highest grade was selected during the audit.

- 2 discrepancies were recoded from a known grade to an unknown grade. 1 case was recoded from moderately differentiated (2) to unknown (9) when the biopsy was taken from a metastatic site, and 1 was recoded from moderately differentiated (2) to unknown (9) when

the pathology report only gave a single area of grade and not a Gleason score due to inadequate tissue.

- 2 discrepancies were recoded from an unknown grade to a known grade. Both of these cases were recoded due to pathology report documentation of grade available in the medical record.

- 1 discrepancy was recoded from poorly differentiated (3) to moderately differentiated (2) when the radiation oncology chart noted the Gleason score of 7.

Date of Diagnosis (mm/dd<=30 days) had 5 discrepancies (83.3 percent) in the high-caseload facilities and 1 discrepancy (16.7 percent) in the medium-caseload facilities. There were no discrepancies in the low-caseload facilities for this data element.

- 5 discrepancies were recoded to an earlier diagnosis date than submitted. All cases involved bone scans, history and physical reports, and surgical orders that documented an earlier date of diagnosis.

- 1 discrepancy was recoded from a nonspecific date, 10/99/2000, to a more specific date upon review of a biopsy report found in the radiation oncology chart.

VIII. CONCLUSION AND RECOMMENDATIONS

ACR had a case completeness rate of 97.5 percent for the four cancer sites audited, which well exceeds the NPCR standard. ACR should be commended for this excellent result and the effort it obviously puts forth in complete case capture. The overall data accuracy rate was 93.4 percent.

ACR's casefinding procedures have already produced an outstanding result and should be continued. Implementation of recommended procedures in this report will help ACR to improve its data quality results. ACR is encouraged to continue conducting visual editing to improve data quality in the State, in addition to reviewing basic abstracting principles.

The TAA auditors noted a few issues that affect the completeness and quality of the cancer data collected at ACR, as discussed below:

1) *Stage at Diagnosis* discrepancies (57) represented 22.4 percent of all discrepancies. 35.1 percent of the *Stage at Diagnosis* discrepancies were identified in lung and bronchus cancers. A review of lung anatomy, methods of metastatic dissemination, diagnostic language, and an in-depth discussion of the differences between SEER Summary Stage and AJCC TNM (Tumor Node and Metastasis) staging rules for lung primaries would help reduce these discrepancies. Staging discrepancies in colon and rectum cancer cases (18) accounted for 31.6 percent of all staging discrepancies, and female breast cancer cases (15) accounted for 26.3 percent.

2) *Subsite* discrepancies (43) accounted for 16.9 percent of all discrepancies. 39.5 percent of those discrepancies were found in female breast cancer cases, and 37.2 percent were found in lung and bronchus cancer cases. A review of the breast clock and the anatomy of the lung would be helpful in reducing the discrepancies in this data element.

3) *Date of Diagnosis* discrepancies (42) constituted 16.5 percent of all the discrepancies. 45.2 percent of those discrepancies were found in lung and bronchus cancer cases, and 23.8 percent were discovered in prostate cancer cases. 23.8 percent of the diagnosis date discrepancies were considered major errors (more than 30 days discrepant). Coding the *Date of Diagnosis* properly is important because it directly affects incidence rates, survival rates, and the determination of the first course of treatment or subsequent treatment. Misinterpretation of diagnostic language affected this field, and reviewing the diagnostic terms will be beneficial to ACR. Close attention to all dictated reports, pathology reports, radiology reports, and visit notes would also be helpful.

4) *Grade* discrepancies (40) represented 15.7 percent of all discrepancies. 35.0 percent occurred in lung and bronchus cancer cases, 30.0 percent in female breast cancer cases, 20.0 percent in prostate cancer cases, and 15.0 percent in colon and rectum cancer cases. Most of the discrepancies occurred due to application of the SBR grading scale when that scale was not documented in the pathology report, use of the lowest grade when multiple grades or grading terms were documented, inattention to the pathology report, or use of the grade from a metastatic site. A review of the ROADS manual, conversion tables, and ACR reporting manual will help to reduce these discrepancies.

5) *Histology* discrepancies (27) represented 10.6 percent of all discrepancies. 37.0 percent of these discrepancies occurred in the lung and bronchus cancer cases, 33.3 percent in female breast cancer cases, 25.9 percent in colon and rectum cancer cases, and 3.7 percent in prostate cancer cases. Reviews for the tumor registrars and registry personnel, with a special focus on multiple histology coding rules and coding of polyps, will help reduce discrepancies in this data element. Attention to all available medical records and documentation, including dictated notes, will help to reduce the number of these discrepancies.

In summary, our recommendations are as follows:

1) Provide a review of basic abstracting practices with a focus on:

   i) Interpretation of the breast clock diagram

   ii) Interpretation of grade conversion from both Gleason's and Scarff Bloom-Richardson grading schemes

   iii) Interpretation of coding multiple histology tumors and polyps

   iv) Determination of the proper information to use in Date of Diagnosis, including reportable terminology and clinical diagnosis versus pathologic confirmation

v) Determination of the SEER Summary Stage and conversion of Tumor, Nodes, and Metastasis (TNM) to the SEER Summary Stage, including a review of the staging timeframe

vi) Use of all dictated reports, such as history and physicals, radiation oncology dictations, and CT scan reports, with careful attention to dates and diagnostic language

vii) Special attention to abstracting and anatomy of female breast, colon and rectum, and lung and bronchus.

2) Strengthen enforcement of the policy on text documentation necessary for quality control procedures

3) Develop a policy outlining the hierarchy of sources to be used in determining the proper Subsite

IX. AUDIT TEAM

The following individuals participated in the ACR audit:

ORC Macro TAA:

1) Lynn Khoo, M.D., M.P.H.
   Principal Investigator
2) Eathell B. Lewis, M.S.
   Program Manager
3) Qiming He, Ph.D.
   Statistician
4) Candi Cain, CTR
   Quality Assurance Specialist
5) Brenda Lange, CTR
   Quality Assurance Specialist
6) Aung Oakkar
   Quality Assurance Assistant
7) Emily Wuerker
   Editor

ACR:

1) Georgia Armenta Yee, BSW, CTR
2) Dina Hudson
3) Brenda Smith

CDC:

1) Faye Floyd, MA. Ed.
2) Kimberly Cantrell
3) Leah Simpson, M.B.A.

The ACR audit was conducted by Candi Cain and Brenda Lange of ORC Macro from September 15 through September 26, 2003. All members of the audit team are trained professionals in the areas of cancer registry operations and management.

APPENDIX

TABLES AND CHARTS

Table 3. Results of the Reconciliation of the Audit Discrepancies (by Major and Minor Errors)
Female Breast, Colon and Rectum, Lung and Bronchus, and Prostate Cancers Combined

NO. OF FACILITIES BY CASELOAD AND REGISTRY STATUS	Records Reabstracted	Major Errors											Minor Errors				Total Errors
		Site*	Dx Date (yy or mm/dd >30 days)**	DOB (yy or mm/dd >30 days)**	Stage	Histol	Race	State	Seq	Lat	Gender	Beh	Subsite	Dx Date (mm/dd <=30 days)**	DOB (mm/dd <=30 days)**	Grade	
High/Registry 4 facilities	132	3	6	0	19	10	4	0	1	5	0	2	19	15	0	19	103
Medium/Registry 3 facilities	99	2	2	0	24	9	5	1	3	7	0	0	9	11	1	14	88
Low/Registry 2 facilities	66	2	2	1	14	8	1	1	1	1	0	2	15	6	2	7	63
Totals	**297**	**7**	**10**	**1**	**57**	**27**	**10**	**2**	**5**	**13**	**0**	**4**	**43**	**32**	**3**	**40**	**254**
ACR Error Rate[1] (%)		2.4	3.4	0.3	19.2	9.1	3.4	0.7	1.7	4.4	0.0	1.3	14.5	10.8	1.0	13.5	6.6
NPCR Error Rate (1993–1999) Mean (%)		3.8	3.3	0.6	10.5	8.7	2.1	0.2	2.6	1.4	0.2	0.7	12.6	8.9	0.9	8.8	5.2
Median (%)		3.8	3.0	0.7	9.8	7.4	1.3	0.0	1.4	1.0	0.0	0.7	10.8	8.4	1.0	8.6	4.8
Range (%)		0.0-10.0	0.0-7.1	0.0-1.4	0.0-34.2	0.3-28.0	0.0-15.2	0.0-1.3	0.0-12.8	0.3-3.4	0.0-1.0	0.0-2.4	1.3-51.0	0.3-20.8	0.0-2.7	0.0-21.0	1.3-13.0
SEER Error Rate (1997) (%)		7.1	0.7	N/A	N/A	4.3	1.6	0.2	0.7	0.9	0.2	0.5	N/A	6.4	N/A	4.3	4.4

Key: *Site=Primary Site (in SEER, Site=Primary Site and Subsite combined)
** Only NPCR 1998 and 1999 are included for Date of Diagnosis and Date of Birth.
Dx Date=Date of Diagnosis
DOB=Date of Birth
Stage=SEER Summary Stage
Histol=Histology
Seq=Sequence Number
Lat=Laterality
Beh=Behavior
[1]Error Rate (%)=Number of Discrepancies x 100/Total Data Elements
Total Data Elements=Number of Records Reabstracted x Number of Data Elements Reviewed

Table 3A. Results of the Reconciliation of the Audit Discrepancies (by Major and Minor Errors)
Female Breast Cancer

NO. OF FACILITIES BY CASELOAD AND REGISTRY STATUS	Records Reabstracted	Major Errors											Minor Errors				Total Errors
		Site*	Dx Date (yy or mm/dd >30 days)**	DOB (yy or mm/dd >30 days)**	Stage	Histol	Race	State	Seq	Lat	Gender	Beh	Subsite	Dx Date (mm/dd <=30 days)**	DOB (mm/dd <=30 days)**	Grade	
High/Registry 4 facilities	38	0	2	0	2	4	1	0	0	0	0	1	8	3	0	4	25
Medium/Registry 3 facilities	32	0	1	0	11	1	3	1	1	0	0	0	5	2	1	5	31
Low/Registry 2 facilities	18	0	0	1	2	4	0	0	0	0	0	1	4	0	0	3	15
Totals	**88**	**0**	**3**	**1**	**15**	**9**	**4**	**1**	**1**	**0**	**0**	**2**	**17**	**5**	**1**	**12**	**71**
ACR Error Rate[1] (%)		0.0	3.4	1.1	17.0	10.2	4.5	1.1	1.1	0.0	0.0	2.3	19.3	5.7	1.1	13.6	6.2
NPCR Error Rate (1993–1999) Mean (%)		4.4	2.4	0.3	7.1	9.9	2.2	0.3	2.5	1.8	0.1	1.4	23.7	8.0	0.8	9.3	4.3
Median (%)		5.0	1.8	0.0	5.8	9.1	1.2	0.0	2.0	0.9	0.0	1.0	25.0	7.8	0.3	9.3	3.3
Range (%)		0.0-10.0	0.0-7.6	0.0-2.0	0.0-23.3	0.7-26.9	0.0-16.5	0.0-2.6	0.0-12.6	0.0-32.9	0.0-0.9	0.0-5.0	2.6-47.6	0.0-17.3	0.0-2.4	0.7-26.9	1.1-11.4
SEER Error Rate (1997) (%)		16.2	N/A	N/A	N/A	6.8	1.5	0.1	0.4	0.5	0.2	0.4	N/A	N/A	N/A	5.9	4.4

Key: *Site=Primary Site (in SEER, Site=Primary Site and Subsite combined)
** Only NPCR 1998 and 1999 are included for Date of Diagnosis and Date of Birth.
Dx Date=Date of Diagnosis
DOB=Date of Birth
Stage=SEER Summary Stage
Histol=Histology
Seq=Sequence Number
Lat=Laterality
Beh=Behavior
[1]Error Rate (%)=Number of Discrepancies x 100/Total Data Elements
Total Data Elements=Number of Records Reabstracted x Number of Data Elements Reviewed

Table 3B. Results of the Reconciliation of the Audit Discrepancies (by Major and Minor Errors)
Colon and Rectum Cancer

NO. OF FACILITIES BY CASELOAD AND REGISTRY STATUS	Records Reabstracted	Major Errors											Minor Errors				Total Errors
		Site*	Dx Date (yy or mm/dd >30 days)**	DOB (yy or mm/dd >30 days)**	Stage	Histol	Race	State	Seq	Lat	Gender	Beh	Subsite	Dx Date (mm/dd <=30 days)**	DOB (mm/dd <=30 days)**	Grade	
High/Registry 4 facilities	34	2	0	0	7	3	1	0	0	0	0	1	6	2	0	5	27
Medium/Registry 3 facilities	15	2	0	0	5	3	0	0	1	2	0	0	0	2	0	1	16
Low/Registry 2 facilities	14	1	0	0	6	1	0	1	0	0	0	1	4	1	0	0	15
Totals	**63**	**5**	**0**	**0**	**18**	**7**	**1**	**1**	**1**	**2**	**0**	**2**	**10**	**5**	**0**	**6**	**58**
ACR Error Rate[1] (%)		7.9	0.0	0.0	28.6	11.1	1.6	1.6	1.6	3.2	0.0	3.2	15.9	7.9	0.0	9.5	7.1
NPCR Error Rate (1993–1999) Mean (%)		4.2	1.3	0.6	12.2	10.8	2.4	0.0	2.5	0.2	0.3	0.6	6.8	7.7	0.8	6.3	3.6
Median (%)		4.4	1.0	0.0	11.3	9.9	1.2	0.0	0.3	0.0	0.0	0.0	6.2	5.0	0.0	5.0	3.5
Range (%)		0.0-10.8	0.0-4.0	0.0-2.0	0.0-36.0	0.0-28.8	0.0-16.9	0.0	0.0-16.9	0.0-1.8	0.0-1.8	0.0-2.0	0.0-18.7	0.0-28.0	0.0-5.9	0.0-23.0	0.3-10.2
SEER Error Rate (1997) (%)		4.5	N/A	N/A	N/A	4.4	1.8	0.2	1.5	0.3	0.6	1.7	N/A	N/A	N/A	2.0	3.6

Key: *Site=Primary Site (in SEER, Site=Primary Site and Subsite combined)
** Only NPCR 1998 and 1999 are included for Date of Diagnosis and Date of Birth.
Dx Date=Date of Diagnosis
DOB=Date of Birth
Stage=SEER Summary Stage
Histol=Histology
Seq=Sequence Number
Lat=Laterality
Beh=Behavior
[1]Error Rate (%)=Number of Discrepancies x 100/Total Data Elements
Total Data Elements=Number of Records Reabstracted x Number of Data Elements Reviewed

Table 3C. Results of the Reconciliation of the Audit Discrepancies (by Major and Minor Errors)
Lung and Bronchus Cancer

NO. OF FACILITIES BY CASELOAD AND REGISTRY STATUS	Records Reabstracted	Major Errors											Minor Errors				Total Errors
		Site*	Dx Date (yy or mm/dd >30 days)**	DOB (yy or mm/dd >30 days)**	Stage	Histol	Race	State	Seq	Lat	Gender	Beh	Subsite	Dx Date (mm/dd <=30 days)**	DOB (mm/dd <=30 days)**	Grade	
High/Registry 4 facilities	29	1	1	0	8	3	0	0	1	5	0	0	5	5	0	4	33
Medium/Registry 3 facilities	28	0	1	0	6	4	1	0	1	3	0	0	4	6	0	6	32
Low/Registry 2 facilities	21	1	1	0	6	3	1	0	1	1	0	0	7	5	2	4	32
Totals	**78**	**2**	**3**	**0**	**20**	**10**	**2**	**0**	**3**	**9**	**0**	**0**	**16**	**16**	**2**	**14**	**97**
ACR Error Rate[1] (%)		2.6	3.8	0.0	25.6	12.8	2.6	0.0	3.8	11.5	0.0	0.0	20.5	20.5	2.6	17.9	9.6
NPCR Error Rate (1993–1999) Mean (%)		4.2	3.7	0.5	13.6	9.9	1.7	0.4	3.2	3.9	0.3	0.2	11.1	10.8	0.8	8.8	4.7
Median (%)		4.5	2.4	0.0	12.7	8.7	1.2	0.0	2.1	2.5	0.0	0.0	10.1	6.1	0.0	8.3	4.6
Range (%)		0.0-11.5	0.0-8.1	0.0-1.8	0.3-34.3	1.0-28.4	0.0-12.3	0.0-2.2	0.0-10.0	0.4-10.9	0.0-1.6	0.0-1.6	0.0-29.0	0.0-30.2	0.0-6.4	0.7-21.0	0.5-11.0
SEER Error Rate (1997) (%)		7.8	N/A	N/A	N/A	5.3	1.2	0.4	0.9	2.7	0.3	0.0	N/A	N/A	N/A	4.2	5.6

Key: *Site=Primary Site (in SEER, Site=Primary Site and Subsite combined)
** Only NPCR 1998 and 1999 are included for Date of Diagnosis and Date of Birth.
Dx Date=Date of Diagnosis
DOB=Date of Birth
Stage=SEER Summary Stage
Histol=Histology
Seq=Sequence Number
Lat=Laterality
Beh=Behavior
[1]Error Rate (%)=Number of Discrepancies x 100/Total Data Elements
Total Data Elements=Number of Records Reabstracted x Number of Data Elements Reviewed

Table 3D. Results of the Reconciliation of the Audit Discrepancies (by Major and Minor Errors)
Prostate Cancer

NO. OF FACILITIES BY CASELOAD AND REGISTRY STATUS	Records Reabstracted	Major Errors											Minor Errors				Total Errors
		Site*	Dx Date (yy or mm/dd >30 days)**	DOB (yy or mm/dd >30 days)**	Stage	Histol	Race	State	Seq	Lat	Gender	Beh	Subsite	Dx Date (mm/dd <=30 days)**	DOB (mm/dd <=30 days)**	Grade	
High/Registry 4 facilities	31	0	3	0	2	0	2	0	0	0	0	0	0	5	0	6	18
Medium/Registry 3 facilities	24	0	0	0	2	1	1	0	0	2	0	0	0	1	0	2	9
Low/Registry 2 facilities	13	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Totals	**68**	**0**	**4**	**0**	**4**	**1**	**3**	**0**	**0**	**2**	**0**	**0**	**0**	**6**	**0**	**8**	**28**
ACR Error Rate[1] (%)		0.0	5.9	0.0	5.9	1.5	4.4	0.0	0.0	2.9	0.0	0.0	0.0	8.8	0.0	11.8	3.2
NPCR Error Rate (1993–1999) Mean (%)		2.6	3.4	0.6	9.2	2.0	2.4	0.2	1.8	0.5	0.1	0.0	0.0	3.3	0.8	11.3	1.9
Median (%)		2.2	2.5	0.0	8.1	1.4	1.3	0.0	0.3	0.0	0.0	0.0	0.0	3.4	0.3	9.8	1.4
Range (%)		0.0-8.4	0.0-9.5	0.0-5.3	0.0-35.7	0.0-18.0	0.0-21.7	0.0-2.7	0.0-12.3	0.0-2.7	0.0-1.3	0.0	0.0-1.2	0.0-8.8	0.0-5.3	0.0-28.8	0.0-6.7
SEER Error Rate (1997) (%)		0.0	N/A	N/A	N/A	0.5	2.0	0.1	0.1	0.1	0.0	0.0	N/A	N/A	N/A	5.1	4.1

Key: *Site=Primary Site (in SEER, Site=Primary Site and Subsite combined)
** Only NPCR 1998 and 1999 are included for Date of Diagnosis and Date of Birth.
Dx Date=Date of Diagnosis
DOB=Date of Birth
Stage=SEER Summary Stage
Histol=Histology
Seq=Sequence Number
Lat=Laterality
Beh=Behavior
[1]Error Rate (%)=Number of Discrepancies x 100/Total Data Elements
Total Data Elements=Number of Records Reabstracted x Number of Data Elements Reviewed

Table 4. Results of the Reconciliation of the Audit Discrepancies (All Errors)

Female Breast, Colon and Rectum, Lung and Bronchus, and Prostate Cancers Combined

NO. OF FACILITIES BY CASELOAD AND REGISTRY STATUS	Records Reabstracted	Site*	Dx Date**	DOB**	Stage	Histol	Race	State	Seq	Lat	Gender	Beh	Subsite	Grade	Total Errors
High/Registry 4 facilities	132	3	21	0	19	10	4	0	1	5	0	2	19	19	103
Medium/Registry 3 facilities	99	2	13	1	24	9	5	1	3	7	0	0	9	14	88
Low/Registry 2 facilities	66	2	8	3	14	8	1	1	1	1	0	2	15	7	63
Totals	**297**	**7**	**42**	**4**	**57**	**27**	**10**	**2**	**5**	**13**	**0**	**4**	**43**	**40**	**254**
ACR Error Rate[1] (%)		2.4	14.1	1.3	19.2	9.1	3.4	0.7	1.7	4.4	0.0	1.3	14.5	13.5	6.6
NPCR Error Rate (1993–1999) Mean (%)		3.8	12.2	1.5	10.5	8.7	2.1	0.2	2.6	1.4	0.2	0.7	12.6	8.8	5.2
Median (%)		3.8	11.1	1.3	9.8	7.4	1.3	0.0	1.4	1.0	0.0	0.7	10.8	8.6	4.8
Range (%)		0.0-10.0	0.3-26.6	0.3-3.0	0.0-34.2	0.3-28.0	0.0-15.2	0.0-1.3	0.0-12.8	0.3-3.4	0.0-1.0	0.0-2.4	1.3-51.0	0.0-21.0	1.3-13.0
SEER Error Rate (1997) (%)		7.1	7.1	0.9	N/A	4.3	1.6	0.2	0.7	0.9	0.2	0.5	N/A	4.3	4.4

Key: *Site=Primary Site (in SEER, Site=Primary Site and Subsite combined)
** Only NPCR 1998 and 1999 are included for Date of Diagnosis and Date of Birth.
Dx Date=Date of Diagnosis
DOB=Date of Birth
Stage=SEER Summary Stage
Histol=Histology
Seq=Sequence Number
Lat=Laterality
Beh=Behavior
[1]Error Rate (%)=Number of Discrepancies x 100/Total Data Elements
Total Data Elements=Number of Records Reabstracted x Number of Data Elements Reviewed

Table 4A. Results of the Reconciliation of the Audit Discrepancies (All Errors)
Female Breast Cancer

NO. OF FACILITIES BY CASELOAD AND REGISTRY STATUS	Records Reabstracted	Site*	Dx Date**	DOB**	Stage	Histol	Race	State	Seq	Lat	Gender	Beh	Subsite	Grade	Total Errors
High/Registry 4 facilities	38	0	5	0	2	4	1	0	0	0	0	1	8	4	25
Medium/Registry 3 facilities	32	0	3	1	11	1	3	1	1	0	0	0	5	5	31
Low/Registry 2 facilities	18	0	0	1	2	4	0	0	0	0	0	1	4	3	15
Totals	**88**	**0**	**8**	**2**	**15**	**9**	**4**	**1**	**1**	**0**	**0**	**2**	**17**	**12**	**71**
ACR Error Rate[1] (%)		0.0	9.1	2.3	17.0	10.2	4.5	1.1	1.1	0.0	0.0	2.3	19.3	13.6	6.2
NPCR Error Rate (1993–1999) Mean (%)		4.4	10.4	1.0	7.1	9.9	2.2	0.3	2.5	1.8	0.1	1.4	23.7	9.3	4.3
Median (%)		5.0	8.9	1.0	5.8	9.1	1.2	0.0	2.0	0.9	0.0	1.0	25.0	9.3	3.3
Range (%)		0.0-10.0	0.0-24.4	0.0-4.0	0.0-23.3	0.7-26.9	0.0-16.5	0.0-2.6	0.0-12.6	0.0-32.9	0.0-0.9	0.0-5.0	2.6-47.6	0.7-26.9	1.1-11.4
SEER Error Rate (1997) (%)		16.2	6.2	0.8	N/A	6.8	1.5	0.1	0.4	0.5	0.0	0.4	N/A	5.9	4.4

Key: *Site=Primary Site (in SEER, Site=Primary Site and Subsite combined)
** Only NPCR 1998 and 1999 are included for Date of Diagnosis and Date of Birth.
Dx Date=Date of Diagnosis
DOB=Date of Birth
Stage=SEER Summary Stage
Histol=Histology
Seq=Sequence Number
Lat=Laterality
Beh=Behavior
[1]Error Rate (%)=Number of Discrepancies x 100/Total Data Elements
Total Data Elements=Number of Records Reabstracted x Number of Data Elements Reviewed

Table 4B. Results of the Reconciliation of the Audit Discrepancies (All Errors)
Colon and Rectum Cancer

NO. OF FACILITIES BY CASELOAD AND REGISTRY STATUS	Records Reabstracted	Site*	Dx Date**	DOB**	Stage	Histol	Race	State	Seq	Lat	Gender	Beh	Subsite	Grade	Total Errors
High/Registry 4 facilities	34	2	2	0	7	3	1	0	0	0	0	1	6	5	27
Medium/Registry 3 facilities	15	2	2	0	5	3	0	0	1	2	0	0	0	1	16
Low/Registry 2 facilities	14	1	1	0	6	1	0	1	0	0	0	1	4	0	15
Totals	**63**	**5**	**5**	**0**	**18**	**7**	**1**	**1**	**1**	**2**	**0**	**2**	**10**	**6**	**58**
ACR Error Rate[1] (%)		7.9	7.9	0.0	28.6	11.1	1.6	1.6	1.6	3.2	0.0	3.2	15.9	9.5	7.1
NPCR Error Rate (1993–1999) Mean (%)		4.2	8.9	1.4	12.2	10.8	2.4	0.0	2.5	0.2	0.3	0.6	6.8	6.3	3.6
Median (%)		4.4	8.0	1.3	11.3	9.9	1.2	0.0	0.3	0.0	0.0	0.0	6.2	5.0	3.5
Range (%)		0.0-10.8	0.0-32.0	0.0-5.9	0.0-36.0	0.0-28.8	0.0-16.9	0.0	0.0-16.9	0.0-1.8	0.0-1.8	0.0-2.0	0.0-18.7	0.0-23.0	0.3-10.2
SEER Error Rate (1997) (%)		4.5	6.8	0.8	N/A	4.4	1.8	0.2	1.5	0.3	0.6	1.7	N/A	2.0	3.6

Key: *Site=Primary Site (in SEER, Site=Primary Site and Subsite combined)
** Only NPCR 1998 and 1999 are included for Date of Diagnosis and Date of Birth.
Dx Date=Date of Diagnosis
DOB=Date of Birth
Stage=SEER Summary Stage
Histol=Histology
Seq=Sequence Number
Lat=Laterality
Beh=Behavior
[1]Error Rate (%)=Number of Discrepancies x 100/Total Data Elements
Total Data Elements=Number of Records Reabstracted x Number of Data Elements Reviewed

Table 4C. Results of the Reconciliation of the Audit Discrepancies (All Errors)
Lung and Bronchus Cancer

NO. OF FACILITIES BY CASELOAD AND REGISTRY STATUS	Records Reabstracted	Site*	Dx Date**	DOB**	Stage	Histol	Race	State	Seq	Lat	Gender	Beh	Subsite	Grade	Total Errors
High/Registry 4 facilities	29	1	6	0	8	3	0	0	1	5	0	0	5	4	33
Medium/Registry 3 facilities	28	0	7	0	6	4	1	0	1	3	0	0	4	6	32
Low/Registry 2 facilities	21	1	6	2	6	3	1	0	1	1	0	0	7	4	32
Totals	**78**	**2**	**19**	**2**	**20**	**10**	**2**	**0**	**3**	**9**	**0**	**0**	**16**	**14**	**97**
ACR Error Rate[1] (%)		2.6	24.4	2.6	25.6	12.8	2.6	0.0	3.8	11.5	0.0	0.0	20.5	17.9	9.6
NPCR Error Rate (1993–1999) Mean (%)		4.2	14.6	1.1	13.6	9.9	1.7	0.4	3.2	3.9	0.3	0.2	11.1	8.8	4.7
Median (%)		4.5	12.8	0.4	12.7	8.7	1.2	0.0	2.1	2.5	0.0	0.0	10.1	8.3	4.6
Range (%)		0.0-11.5	0.4-38.4	0.0-4.3	0.3-34.3	1.0-28.4	0.0-12.3	0.0-2.2	0.0-10.0	0.4-10.9	0.0-1.6	0.0-1.6	0.0-29.0	0.7-21.0	0.5-11.0
SEER Error Rate (1997) (%)		7.8	9.7	1.0	N/A	5.3	1.2	0.4	0.9	2.7	0.3	0.0	7.8*	4.2	5.6

Key: *Site=Primary Site (in SEER, Site=Primary Site and Subsite combined)
** Only NPCR 1998 and 1999 are included for Date of Diagnosis and Date of Birth.
Dx Date=Date of Diagnosis
DOB=Date of Birth
Stage=SEER Summary Stage
Histol=Histology
Seq=Sequence Number
Lat=Laterality
Beh=Behavior
[1]Error Rate (%)=Number of Discrepancies x 100/Total Data Elements
Total Data Elements=Number of Records Reabstracted x Number of Data Elements Reviewed

Table 4D. Results of the Reconciliation of the Audit Discrepancies (All Errors)
Prostate Cancer

NO. OF FACILITIES BY CASELOAD AND REGISTRY STATUS	Records Reabstracted	Site*	Dx Date**	DOB**	Stage	Histol	Race	State	Seq	Lat	Gender	Beh	Subsite	Grade	Total Errors
High/Registry 3 facilities	31	0	8	0	2	0	2	0	0	0	0	0	0	6	18
Medium/Registry 3 facilities	24	0	1	0	2	1	1	0	0	2	0	0	0	2	9
Low/Registry 2 facilities	13	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Totals	**68**	**0**	**10**	**0**	**4**	**1**	**3**	**0**	**0**	**2**	**0**	**0**	**0**	**8**	**28**
ACR Error Rate[1] (%)		0.0	14.7	0.0	5.9	1.5	4.4	0.0	0.0	2.9	0.0	0.0	0.0	11.8	3.2
NPCR Error Rate (1993–1999) Mean (%)		2.6	6.7	1.4	9.2	2.0	2.4	0.2	1.8	0.5	0.1	0.0	0.0	11.3	1.9
Median (%)		2.2	5.3	1.0	8.1	1.4	1.3	0.0	0.3	0.0	0.0	0.0	0.0	9.8	1.4
Range (%)		0.0-8.4	0.0-17.5	0.0-10.5	0.0-35.7	0.0-18.0	0.0-21.7	0.0-2.7	0.0-12.3	0.0-2.7	0.0-1.3	0.0	0.0-1.2	0.0-28.8	0.0-6.7
SEER Error Rate (1997) (%)		0.0	5.8	0.7	N/A	0.5	2.0	0.1	0.1	0.1	0.0	0.0	N/A	5.1	4.1

Key: *Site=Primary Site (in SEER, Site=Primary Site and Subsite combined)
**Only NPCR 1998 and 1999 are included for Date of Diagnosis and Date of Birth.
Dx Date=Date of Diagnosis
DOB=Date of Birth
Stage=SEER Summary Stage
Histol=Histology
Seq=Sequence Number
Lat=Laterality
Beh=Behavior
[1]Error Rate (%)=Number of Discrepancies x 100/Total Data Elements
Total Data Elements=Number of Records Reabstracted x Number of Data Elements Reviewed

Figure 1. Data Discrepancy Rates by Primary Site for Diagnosis Year 2000—Arizona Cancer Registry

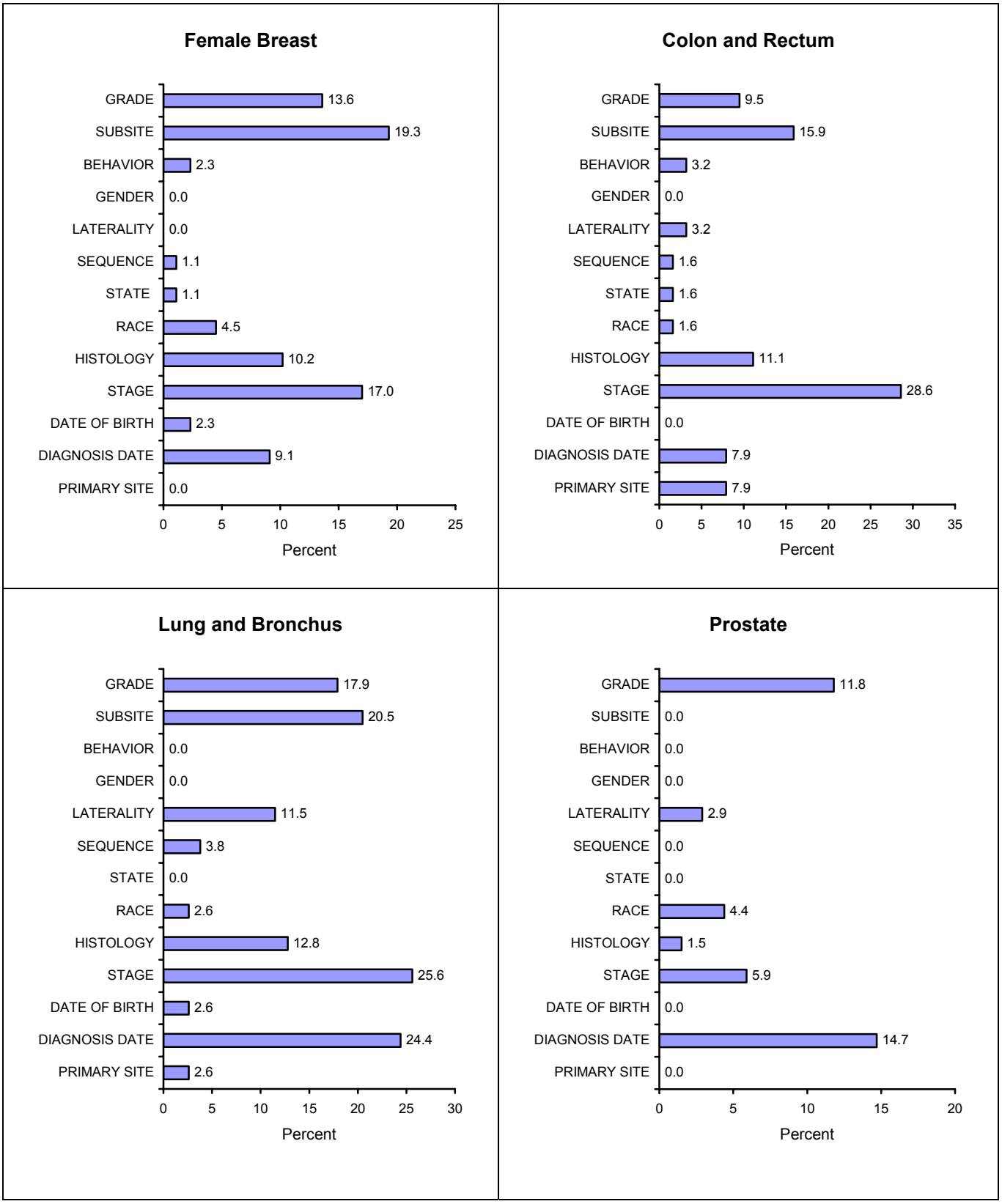
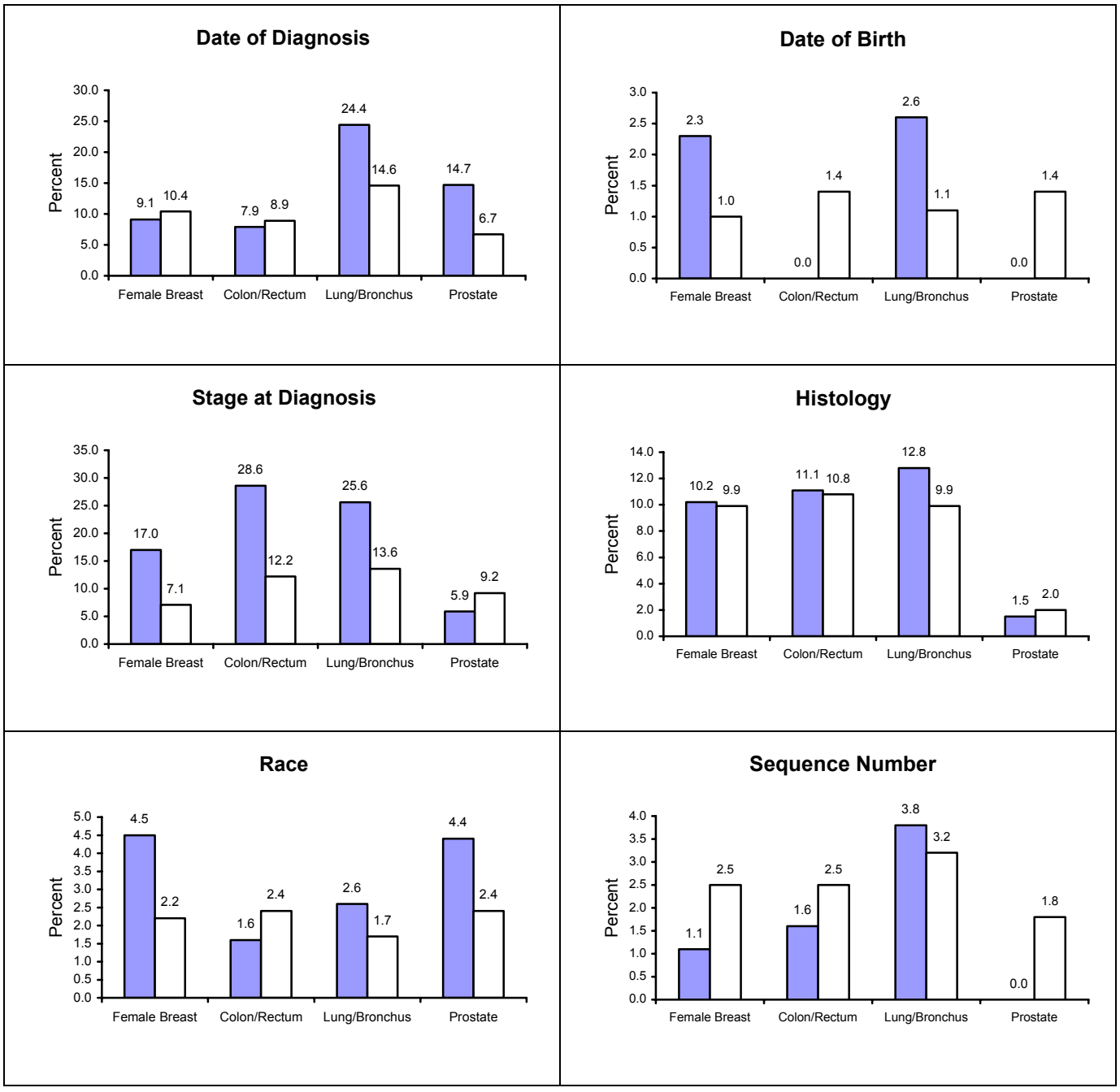


Figure 2. Comparison of Data Discrepancy Rate of Selected Data Element for Female Breast, Colon and Rectum, Lung and Bronchus, and Prostate Cancers—Arizona Cancer Registry (2000) and NPCR (1993–1999) *



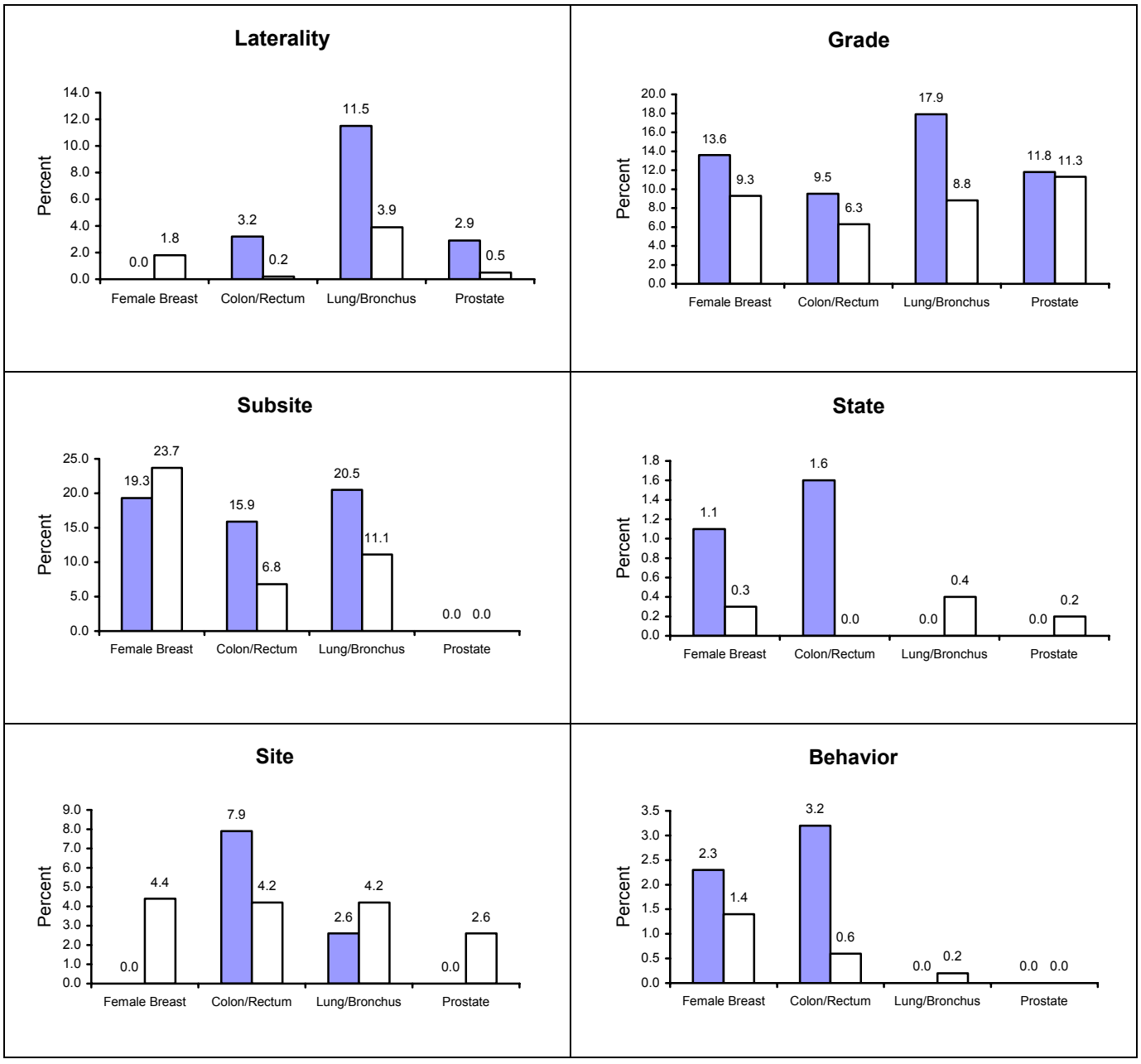
* Only NPCR 1998 and 1999 are included for Date of Diagnosis and Date of Birth.

Legend:

Arizona Cancer Registry (2000)

NPCR (1993–1999)*

Figure 2, cont. Comparison of Data Discrepancy Rate of Selected Data Element for Female Breast, Colon and Rectum, Lung and Bronchus, and Prostate Cancers—Arizona Cancer Registry (2000) and NPCR (1993–1999)*



* Only NPCR 1998 and 1999 are included for Date of Diagnosis and Date of Birth.

Legend:


Arizona Cancer Registry (2000)


NPCR (1993–1999)*

APPENDIX

CALCULATING VARIANCES AND STANDARD ERRORS

Calculating Variances (V) and Standard Errors (SE)

The variance of a stratified estimate of the proportion is: V(p) = sum {W(h)**2}V(h).

Here, V(h) is the variance of p(h), and can be expressed as:

{(N(h)-n(h))/n(h)}*{p(h)(1-p(h))/(n(h)-1)}

The number of eligible cases in stratum-h (h=1, 2, 3) is denoted by N(h), the total State caseload (eligible cases) is N = N(1) + N(2) + N(3). Stratum weights, W(h)=N(h)/N, are the share of each stratum in the total caseload.

In both scenarios—completeness and data quality assessment—rates are first calculated within each stratum to generate stratum-level estimates, p(h). The variance of p(h) is designated by V(h), and its square root is the standard error, se(h).

The variance V(h) is calculated as: p(h)*{1-p(h)}/n(h).

The sample hospitals are selected with probabilities proportional to size (PPS) from the sampling frame. Then hospitals are divided into three categories, or strata (high, medium, and low), based on caseload. In each sample hospital, a fixed number of cases are reviewed regardless of the category. This design yields a self-weighting sample in each stratum, i.e., cases are selected with equal probabilities so that sampling weights are equal within each stratum.

First, the total numbers of cases in each caseload facilities qualified for audit are summed. Then, the proportion of cases in each caseload category is calculated based on total number of cases in the sampling frame, generating the weight for each stratum.

The weighted estimate of the proportion is: p = sum W(h)*p(h)

where all the sums are over all three strata, and strata are indexed by "h" (1, 2, 3).

This is the simple random sampling variance for a proportion in the stratum-h sample of size n(h).

Overall estimates are then p = sum W(h)*p(h) (sum over all three strata); their variances are computed as: V(p) = sum {W(h)**2}V(h).

$$SE = \sqrt{\text{Variance}}$$

The standard error, se(p), is the square root of this variance, and provides the basis for confidence intervals.

For case completeness rates, the denominator (n) is the cases reviewed for the number of months reviewed plus number of missed cases per category (stratum).

For data quality assessment, the denominator (n) is the number of records reviewed.

Reference: Cochran, 1977, Chapter 5.

APPENDIX

REFERENCES

References

1. International Classification of Diseases for Oncology, second edition. Ed. Constance Percy, Valerie Van Holten, Calum Muir. Geneva: World Health Organization, 1990.

2. American Joint Commission on Cancer Staging Manual, fifth edition. American Joint Commission on Cancer. Lippincott-Raven Publishers, 1997.

3. Registry Operations and Data Standards (ROADS), volume II. American College of Surgeons. Chicago: American College of Surgeons, January 1998.

4. Summary Staging Guide, April 1977. Surveillance, Epidemiology, and End Results (SEER) Program. Washington, DC: U.S. Department of Health and Human Services. Reprinted July 1986.

5. SEER Program Manual, third edition. Cancer Statistics Branch, National Cancer Institute. Washington, DC: U.S. Department of Health and Human Services, January 1998.

6. SEER Inquiry System (www.seer.cancer.gov). Maintained by the SEER Program, National Cancer Institute. Updated annually.

7. Inquiry and Response System of the American College of Surgeons (http://www.facs.org/dept/cancer/coc/iandr.html). Maintained by the Commission on Cancer. Updated annually.

8. Standards for Cancer Registries, volume II: Data Standards and Data Dictionary, sixth edition, record layout version 9.1. Ed. Dianne Hultstrom. North American Association of Central Cancer Registries, March 2001.

9. Arizona Cancer Registry Data Dictionary.